

Lo studio della razionalità e della ragione pratica, o “razionalità in azione” come la definisce John Searle, occupa un ruolo centrale nella storia intellettuale dell’Occidente. In questo libro, che sostiene una tesi “forte” a favore della razionalità, Searle espone le sei tesi di quello che definisce il “Modello Classico” della razionalità e dimostra perché esse siano tutte false. Egli presenta poi una teoria alternativa del ruolo della razionalità nel pensiero e nell’azione.

Un punto centrale della teoria di Searle è che sono soltanto le azioni irrazionali che tendono ad essere causate da credenze e desideri: per esempio, le azioni di una persona determinate da un’ossessione patologica o dall’uso di sostanze stupefacenti. Nella maggior parte dei casi di azioni razionali c’è un divario (*gap*) tra il desiderio motivante e l’effettiva decisione presa. Il nome tradizionalmente attribuito a questo divario è “libertà della volontà” o “libero arbitrio”. Secondo Searle, qualsiasi attività razionale presuppone il libero arbitrio, per cui la razionalità è possibile soltanto dove è possibile una scelta tra diverse opzioni, sia razionali che irrazionali. Diversamente da altri trattati filosofici, *La razionalità in azione* invita il lettore ad applicare le idee dell’autore alla vita reale. Searle dimostra, ad esempio, che, contrariamente a quanto sostenuto dalla tradizione filosofica, la debolezza della volontà è molto comune. Egli mette inoltre in evidenza l’assurdità della tesi secondo cui il prendere delle decisioni razionali prende sempre avvio di un insieme coerente di desideri. Il prendere decisioni razionali, egli sostiene, ha spesso a che fare con la scelta tra ragioni per l’azione tra loro in conflitto. Gli esseri umani si caratterizzano infatti per la loro capacità di essere razionalmente motivati da ragioni per l’azione indipendenti dal desiderio. Estendendo questa teoria alla razionalità del sé, Searle dimostra come la deliberazione razionale presupponga una nozione irriducibile del sé. Egli sostiene inoltre la tesi secondo la quale per comprendere il libero arbitrio è essenziale la comprensione di come funziona il nostro cervello.

John R. Searle è professore di Filosofia della mente e Filosofia del linguaggio all’Università della California, Berkeley. Tra le sue opere tradotte in italiano ricordiamo *Dell’intenzionalità* (Milano, 1985), *Mente, cervello, intelligenza* (Milano, 1988), *Atti linguistici* (Torino, 1992), e, pubblicati in questa collana, *Il mistero della coscienza* (1998) e *Mente, linguaggio, società* (2000).



John R. Searle

La razionalità nell'azione

“Rationality in Action”

The MIT Press

© 2001 Massachusetts Institute of Technology

Cambridge, Massachusetts London, England

Traduzione:

Eddy Carli

Mario Valentino Bramè

INDICE

Avvertenza sulla collana

Ringraziamenti

Introduzione

1. Il modello classico della razionalità e la sua debolezza
2. La struttura di base dell'intenzionalità, dell'azione e del significato
3. Il divario: il tempo e il sé
4. La struttura logica delle ragioni
5. Alcune caratteristiche speciali della ragione pratica: un forte altruismo come requisito logico
6. In che modo creiamo delle ragioni per l'azione indipendenti dai desideri

Appendice. Ragioni interne ed esterne

7. La debolezza della volontà
8. Perché non c'è una logica deduttiva della ragione pratica
9. La coscienza, la libera azione e il cervello

Avvertenza sulla collana

Le “Jean Nicod Lectures” si svolgono annualmente a Parigi e sono tenute da un autorevole filosofo della mente o da uno studioso di scienze cognitive orientato verso la filosofia. Le conferenze inaugurali del 1993 celebrarono il centenario della nascita del filosofo francese Jean Nicod (1893-1931). Le conferenze sono sostenute dal *Centre National de la Recherche Scientifique* (CNRS) e sono organizzate in cooperazione con la *Fondation Maison des Sciences de l’Homme* (MSH). La collana accoglie i testi delle conferenze o le monografie a essi ispirati.

Jacques Bouveresse, presidente del *Jean Nicod Committee*

François Recanati, segretario del *Jean Nicod Committee* ed editore della collana

Jean Nicod Committee

Mario Borillo

Jean-Pierre Changeux

Jean-Gabriel Ganascia

André Holley

Michel Imbert

Pierre Jacob

Jacques Mehler

Elisabeth Pacherie

Philippe de Rouilhan

Dan Sperber

RINGRAZIAMENTI

Paul Valéry, da qualche parte, ha detto: non si finisce mai una poesia, la si abbandona nella disperazione. Lo stesso si può dire di alcuni lavori filosofici. Mi è capitato più di una volta di avere terminato un libro, di averlo finalmente spedito all'editore e di aver pensato: "se solo potessi riscriverlo dall'inizio, ora che so come dovrebbe essere scritto!" Ebbene, questo libro è stato davvero riscritto da capo. Avevo finito la prima stesura qualche anno fa ed era stato accettato per la pubblicazione, ma poi decisi di riscrivere il tutto. Eliminaí completamente alcuni capitoli, ne aggiunsi degli altri, e riscrissi diverse parti di quelli che erano rimasti. Ora che è dall'editore ho ancora lo stesso pensiero: "se solo potessi riscriverlo dall'inizio..."

In parte a causa dell'argomento molto vasto, ho ricevuto più commenti utili del solito, e sono quindi più in debito del solito nei confronti dei miei studenti e degli altri critici. Questo materiale è stato argomento di seminari che ho tenuto a Berkeley così come di diverse lezioni tenute nel Nord America, in Europa, in Sud America e in Asia. È stato l'argomento di un simposio, all'annuale convegno dedicato a Wittgenstein a Kirchberg, in Austria, nel 2000, e quattro dei capitoli sono stati presentati alle "Nicod Lectures" di Parigi e alle "Tasan Memorial Lectures" di Seul, in Corea. Una prima versione di questo testo, corrispondente a gran parte dei primi sette capitoli, ha ricevuto il premio internazionale "Jovellanos", in Spagna, ed è stato pubblicato in spagnolo con il titolo *Razones Para Actuar* (Ediciones Nobel), nel 2000. Ho un debito speciale di gratitudine nei confronti del mio traduttore spagnolo, Luis Valdes. Il lavoro sul materiale di questo libro ha avuto inizio, in sostanza, più di quindici anni fa, quando Michael Bratman mi invitò a tenere una conferenza sulla ragione pratica. Sono in debito con Michael e gli altri partecipanti alla conferenza per le loro critiche. Sono grato a Chris Cowell per aver preparato l'indice e alle tante persone che hanno letto e commentato le diverse parti del manoscritto, specialmente Robert Audi, Guido Bacciagaluppi, Berit Brogaard-Pedersen, Winston Chiong, Alan Code, Boudewijn de Bruin, Jennifer Hudin, Christine Korsgaard, Josef Moural, Thomas Nagel, Jessica Samuels, Barry Smith, Mariam Thalos, Bernard Williams, Leo Zaibert e, soprattutto, mia moglie, Dagmar Searle, alla quale questo libro è dedicato.

INTRODUZIONE

Questo libro presenta quella che, con alcune riserve, si potrebbe qualificare come una teoria della ragione pratica. La prima riserva è che l'argomento è così vasto e complesso che, all'interno di un libro che ha tale scopo, si possono trattare solo alcuni dei problemi centrali.

Talvolta lavoro meglio quando posso contrapporre la prospettiva che sto presentando a delle prospettive contrarie. La filosofia spesso procede con la discussione e il dibattito. In questo caso, la prospettiva contraria è la concezione della razionalità all'interno della quale sono cresciuto e che penso sia la concezione dominante nella nostra cultura intellettuale. La chiamerò, spero non ingiustamente, il "Modello Classico".

Criticando il modello classico, critico una tradizione molto potente nella filosofia occidentale. In questo libro indicherò alcuni dei suoi limiti e proverò a superarli. Tuttavia può sembrare eccessivamente critico attaccare un modello di razionalità che, per molti aspetti, è corretto e che enfatizza il ruolo della razionalità e dell'intelligenza nel prendere delle decisioni e nella vita in generale, in un'epoca in cui vi sono attacchi sistematici ad un'idea forte di razionalità. Varie forme di relativismo, talvolta sotto l'etichetta di "postmodernismo", hanno attaccato l'idea di razionalità in quanto tale. Si suppone che la razionalità sia essenzialmente opprimente, egemonica, culturalmente relativa, ecc. Perché critico allora una buona teoria della razionalità quando la razionalità in quanto tale è sotto attacco? Sono spaventato da questi attacchi come chiunque altro, ma non mi preoccuperò di replicare perché non credo risultino nemmeno intelligibili. Per esempio, certe volte sono stato sfidato con domande del tipo "qual è la tua argomentazione a sostegno della razionalità?" Ebbene si tratta di una sfida priva di senso, dal momento che la nozione stessa di "argomentazione" presuppone dei criteri di razionalità. Questo libro non è una difesa della razionalità, perché l'idea di una difesa in forma di argomentazione, ragioni ecc. presuppone delle assunzioni di razionalità e quindi la richiesta di tale difesa è priva di senso.

Le assunzioni di razionalità sono universali e radicate nella struttura della mente e nel linguaggio, più precisamente nelle strutture dell'intenzionalità e degli atti linguistici. Si possono descrivere le operazioni di queste assunzioni, come io provo a fare in questo libro, e si possono criticare altre descrizioni simili, come pure faccio, ma la razionalità in quanto tale non richiede né ammette alcuna giustificazione, poiché il pensiero e il linguaggio, e quindi l'argomentazione, presuppongono la razionalità. Si possono discutere in maniera intelligibile le teorie della razionalità, ma non la razionalità.

Questo libro è una discussione che si svolge all'interno degli approcci filosofici tradizionali alla razionalità ed è un tentativo di migliorare la prospettiva dominante della tradizione.

Nelle reazioni alle conferenze che ho tenuto su questi argomenti ho trovato due errori persistenti che molte persone intelligenti compiono riguardo a ciò che essi si aspettano da una teoria della razionalità, e vorrei bloccare tali errori fin dall'inizio.

Per prima cosa, molti credono che una teoria della razionalità dovrebbe fornire loro un algoritmo per poter effettuare delle decisioni razionali. Pensano che non spenderebbero bene il loro denaro se acquistassero un libro sulla razionalità che non fornisca loro un metodo concreto per decidere se divorziare o meno, quali investimenti fare al mercato azionario e quale candidato votare alle prossime elezioni. Per ragioni che sono implicite nell'analisi che esporrò, nessuna teoria della razionalità fornirà un algoritmo per prendere le giuste decisioni. L'obiettivo di una teoria di questo tipo non è quello di dirvi come risolvere delle questioni particolari, ma di spiegare certe caratteristiche strutturali del prendere decisioni razionali. Così come una teoria della verità non vi fornirà un algoritmo per scoprire quali proposizioni sono vere, allo stesso modo una teoria della razionalità non vi fornirà un algoritmo per prendere le decisioni più razionali.

Un secondo errore che le persone commettono riguardo alla razionalità è il supporre che se i criteri di razionalità fossero universali e se noi fossimo degli agenti perfettamente razionali, allora non avremmo alcun disaccordo. Di conseguenza, queste persone suppongono che la persistenza di un disaccordo tra agenti apparentemente istruiti e razionali dimostri che la razionalità è, in qualche modo, relativa alle culture e agli individui. Ma tutto questo è un errore. I criteri della razionalità, come i criteri della verità, sono infatti universalmente validi per tutti gli individui e le culture. Ma una volta assunti dei criteri universali di razionalità e data la deliberazione razionale degli agenti, grandi disaccordi sono ancora possibili, e in effetti inevitabili. Assumiamo dei criteri di razionalità universalmente

validi e universalmente accettati, assumiamo degli agenti perfettamente razionali e che operano con perfetta cognizione di causa: troveremo che un disaccordo razionale sarà ancora presente; questo perché gli agenti razionali potrebbero, per esempio, avere dei valori diversi e inconciliabili, ognuno dei quali potrebbe essere razionalmente accettabile. Uno degli errori più radicati all'interno delle assunzioni del nostro sfondo sociale è l'idea che i conflitti irrisolvibili siano un segnale di comportamento irrazionale o, peggio ancora, che la razionalità stessa sia in questione.

Molti dei problemi discussi in questo libro sono tradizionalmente pensati come parte dell'etica filosofica, nel senso che sono quel tipo di problemi affrontati nei corsi universitari di "teoria etica". Ho molto poco da dire riguardo all'etica o circa le possibili implicazioni della mia teoria per la teoria etica. Non sono sicuro che esista una ben definita branca della filosofia chiamata "teoria etica", ma se ne ammettiamo l'esistenza il suo presupposto necessario è una spiegazione della razionalità nel prendere decisioni e nell'agire. Non si possono discutere in modo intelligente, per esempio, le ragioni etiche dell'azione se prima non si sa che cos'è un'azione e che cos'è una ragione. Così questo libro, sebbene non riguardi direttamente l'etica, ha a che fare con molti dei problemi fondazionali di qualsiasi teoria etica.

Questa indagine è una continuazione dei miei precedenti lavori sui problemi della mente, del linguaggio e della realtà sociale. Ognuno di quei libri deve essere considerato in modo autonomo, ma nello stesso tempo ognuno di essi è parte di una più estesa struttura filosofica di tipo generale. Per consentire a questo libro di reggersi da solo, ho riassunto nel capitolo 2 alcuni elementi essenziali dei miei precedenti lavori che ne aiuteranno la comprensione.

IL MODELLO CLASSICO DELLA RAZIONALITÀ E LA SUA DEBOLEZZA

I. IL PROBLEMA DELLA RAZIONALITÀ

Durante la prima guerra mondiale un famoso psicologo animale, Wolfgang Köhler, che lavorava sull'isola di Tenerife, mostrò che le scimmie erano in grado di prendere decisioni razionali. Effettuando un esperimento tipico, mise una scimmia in un contesto ambientale costituito da una cassa, un bastone e un casco di banane posto in alto, impossibile da raggiungere. Dopo un po' la scimmia capì come prenderle. Portò la cassa sotto le banane, prese il bastone, salì sulla cassa e le tirò giù.¹ Köhler si interessava più della psicologia della *Gestalt* che della razionalità, ma le sue scimmie mostravano una forma di razionalità che si è rivelata paradigmatica nelle nostre teorie. L'idea è che il prendere decisioni razionali riguarda la selezione dei *mezzi* che ci permettono di raggiungere i nostri *scopi*. Gli scopi riguardano interamente ciò che desideriamo. Noi giungiamo alle decisioni costruendo delle situazioni attraverso una sorta di elenco iniziale di scopi desiderati, e la razionalità riguarda interamente il problema di come individuare i mezzi per raggiungere tali scopi.

¹ Wolfgang Köhler, *The Mentality of Apes*, II edition, London: Routledge and Keagan Paul, 1927. Gli animali in questione erano degli scimpanzé, tr. it Wolfgang Köhler, *L'intelligenza nelle scimmie antropoidi*, presentazione e traduzione di Guido Petter, Giunti-Barbera, Firenze 1968.

Non c'è dubbio che le scimmie dimostrino una capacità di prendere decisioni razionali simile a quella dell'uomo. Ma c'è un grandissimo numero di altre decisioni razionali che le scimmie non riescono a contemplare, né potrebbero farlo. Le scimmie potrebbero tentare di capire come ottenere delle banane adesso, ma non potrebbero tentare di capire come ottenere delle banane la prossima settimana. Per gli esseri umani, diversamente dalle scimmie, molte decisioni riguardano l'organizzazione del tempo al di là del presente immediato. Inoltre, le scimmie non riescono a considerare un lungo periodo di tempo culminante con la loro morte. Molte decisioni umane, per la verità gran parte delle decisioni più importanti, come il decidere dove andare a vivere, che tipo di carriera intraprendere, che tipo di famiglia avere, chi sposare, hanno a che fare con la distribuzione del tempo precedente alla morte. La morte, si potrebbe dire, è l'orizzonte della razionalità umana; ma i pensieri che riguardano la morte e l'abilità di pianificare con il pensiero della morte sembra oltrepassare i limiti dell'apparato concettuale delle scimmie. Una seconda differenza tra la razionalità umana e il caso delle scimmie è che gli esseri umani sono spesso spinti a scegliere tra scopi in conflitto e tra loro incompatibili. In alcuni casi questo risulta vero in riferimento alla capacità decisionale degli animali – l'asino di Buridano è un famoso caso ipotetico – ma per la scimmia di Köhler c'erano la cassa, il bastone, le banane o niente. La terza limitazione della scimmia è che essa non può considerare ragioni per l'azione che non siano dipendenti dai suoi desideri. Cioè sembra che il suo desiderio di fare qualcosa con la sedia e il bastone sia motivato solo da un precedente desiderio di mangiare le banane. Ma nel caso degli esseri umani risulta che noi disponiamo di un numero piuttosto grande di ragioni che non sono desideri. Queste ragioni indipendenti dai desideri (*desire-independent reasons*) possono costituire il terreno per i desideri, ma il loro essere ragioni per noi non dipende dal loro essere basate su dei desideri. Questo è un punto interessante e controverso sul quale mi soffermerò in maggiore dettaglio nei prossimi capitoli. Un quarto punto di differenza tra noi stessi e le scimmie è che la scimmia sembra avere solamente una concezione molto limitata di sé stessa, ovvero come quella di un agente razionale in grado di prendere decisioni e capace di assumersi le responsabilità future delle decisioni prese nel presente, o responsabilità nel presente per decisioni prese nel passato. Una quinta differenza, legata alla quarta, è che lo scimpanzé, diversamente dagli uomini, non considera in nessun modo le sue decisioni come espressioni di, o impegni verso, principi generali che si applicano, allo stesso modo, sia a sé stessi che agli altri.

In queste discussioni, si è soliti dire che ciò che manca alla scimmia è il linguaggio. L'idea, apparentemente, è che se solo potessimo riuscire a insegnare alle scimmie i rudimenti della comunicazione linguistica, esse disporrebbero della serie completa dell'apparato delle decisioni razionali e delle responsabilità di cui l'uomo dispone. Ho forti dubbi che le cose stiano così. La semplice capacità di simbolizzazione non è di per sé sufficiente a giustificare l'intera gamma dei processi del pensiero razionale. Gli sforzi per insegnare agli scimpanzé ad usare simboli linguistici hanno portato solamente a dei risultati ambigui. Ma anche quando hanno avuto successo, mi sembra che le modalità di uso dei simboli insegnati intenzionalmente a Washoe, a Lana e ad altri famosi scimpanzé, in contesti sperimentali, si siano dimostrati insufficienti a descrivere le diverse capacità razionali umane che accompagnano alcune speciali caratteristiche delle abilità linguistiche umane. Il punto è che la semplice capacità di simbolizzazione non produce di per sé le diverse caratteristiche della razionalità umana. Ciò che è necessario, come vedremo in queste pagine, è la capacità di disporre di certi tipi di rappresentazioni linguistiche; e per questi tipi di rappresentazioni mi sembra non sia possibile operare una chiara distinzione tra capacità intellettuali, che sono espresse nel segno, e l'uso del segno stesso. Il punto chiave è questo: gli animali possono ingannare ma non possono mentire. L'abilità a mentire è una conseguenza della più profonda abilità umana ad assumere certi tipi di *impegni*, e quegli impegni costituiscono i casi in cui l'animale umano impone intenzionalmente *condizioni di soddisfazione su condizioni di soddisfazione*. Se non capite questo punto, non preoccupatevi; lo spiegherò nei prossimi capitoli.

I problemi filosofici persistenti, come il problema della razionalità, hanno una struttura logica caratteristica: come può essere il caso che p , dato che sembra certo il caso che q , e visto che q apparentemente rende impossibile che p ? Il classico esempio di questo modello è, certamente, il problema del libero arbitrio. Come è possibile che noi compiamo delle azioni libere, posto che ogni evento ha una causa, e che la determinazione causale rende impossibili le azioni libere? La stessa struttura logica si può applicare ad un gran numero di altri problemi. Come può essere che abbiamo una coscienza, dato che siamo composti interamente di parti di materia non cosciente? Lo stesso problema sorge con l'intenzionalità: come può essere che abbiamo degli stati intenzionali – stati che si riferiscono agli oggetti e stati di cose nel mondo che vanno oltre noi stessi – dato che siamo fatti di parti di materia che mancano di intenzionalità? Un problema analogo sorge nello scetticismo: come possiamo conoscere qualcosa, dato che non possiamo mai essere certi di

non stare sognando, di non essere preda di allucinazioni, o di essere ingannati da demoni maligni? E nel caso dell'etica: come possono esserci dei valori nel mondo, dato che il mondo consiste interamente di fatti neutrali dal punto di vista del valore? Una variante della stessa domanda: come possiamo, in generale, sapere che cosa *dovrebbe* accadere dato che la conoscenza riguarda ciò che *effettivamente* è, e che non possiamo mai derivare un'asserzione che riguardi un "che cosa dovrebbe essere" da qualsivoglia insieme di asserzioni che riguardano "ciò che effettivamente è"? Il problema della razionalità, inteso come una variante di questi problemi persistenti, può essere posto nel modo segue. Come può esistere la capacità di prendere decisioni razionali in un mondo dove tutto ciò che esiste accade come risultato di forze causali naturali, brute e cieche?

II. IL MODELLO CLASSICO DELLA RAZIONALITÀ

Nella discussione sulla razionalità delle scimmie, ho sottolineato che nella nostra cultura abbiamo una tradizione specifica nel discutere la razionalità, la ragione pratica, la razionalità in azione. Questa tradizione risale alla tesi di Aristotele secondo cui la deliberazione riguarda sempre i mezzi, e mai i fini,² e prosegue con la famosa posizione di Hume secondo la quale "la ragione è e dovrebbe essere la schiava delle passioni" e con la posizione di Kant secondo la quale "colui che vuole un fine vuole i mezzi". La tradizione riceve la sua più sofisticata formulazione nella contemporanea teoria matematica delle decisioni. La tradizione è senza dubbio unificata ed io non vorrei suggerire che Aristotele, Hume e Kant condividano la stessa concezione della razionalità. Al contrario, vi sono differenze notevoli tra loro. Ma c'è un filo conduttore, e credo che tra i filosofi classici sia Hume a fornire la più chiara definizione di quello a cui mi riferirò come al "Modello Classico". Per lungo tempo ho nutrito dei dubbi riguardo a questa tradizione e dedicherò gran parte di questo primo capitolo ad esporre alcune delle principali caratteristiche, e a fornire una preliminare esposizione dei miei dubbi. Un modo per descrivere il Modello Classico è affermare che esso rappresenta la razionalità umana come una versione più complessa della razionalità delle scimmie.

Quando studiai per la prima volta la teoria matematica della decisione, da studente universitario a Oxford, mi sembrò che essa implicasse un problema ovvio: essa sembra essere la stretta conseguenza di assiomi del tipo che se valuto la mia vita e valuto

² Alan Code mi ha fatto notare che questa attribuzione tradizionale potrebbe tuttavia costituire un fraintendimento della tesi originale di Aristotele.

venticinque centesimi (un quarto di dollaro non è molto, ma è sufficiente a farci scendere dal marciapiede per raccoglierlo, per esempio), deve esserci un qualche motivo per il quale rischierei la mia vita per un quarto di dollaro. Pensavo questo e concludevo che non vi è alcun motivo per il quale metterei a repentaglio la mia vita per un quarto di dollaro e, se ci fosse, non scommetterei la vita di mio figlio per un quarto di dollaro. Così, nel corso degli anni, ho discusso di questo con diversi famosi teorici della decisione, a iniziare da Jimmy Savage ad Ann Arbor, a Isaac Levi a New York, e di solito, dopo circa mezz'ora di discussione, essi arrivavano a questa conclusione: "Ti stai semplicemente lamentando in maniera irrazionale". Ebbene, non ne sono così sicuro. Credo, piuttosto, che fossero loro ad avere dei problemi con la teoria della razionalità. Alcuni anni dopo, i limiti di questa concezione della razionalità mi si presentarono dinanzi con una forte evidenza reale (e questo ha una certa importanza pratica): accadde durante il periodo della guerra del Vietnam, in occasione della mia visita ad un amico, che era un alto ufficiale del Dipartimento della Difesa al Pentagono. Provai a ragionare con lui sulla strategia di guerra che gli Stati Uniti stavano seguendo, in particolare la strategia di bombardare il Vietnam del Nord. Egli era dottore di ricerca in economia matematica. Andò alla lavagna e disegnò le curve della analisi microeconomica tradizionale; e poi disse: "Nel punto in cui queste due curve si intersecano, l'utilità marginale di resistenza è uguale alla non-utilità di essere bombardati. A quel punto si devono arrendere. Tutti noi stiamo assumendo che essi sono razionali. Stiamo assumendo che il nemico sia razionale!"

Capii allora che la situazione era seriamente problematica, non solo per la nostra teoria della razionalità ma anche per la sua applicazione nella prassi. Sembra sia una assunzione da pazzi ritenere che la decisione di affrontare Ho Chi Minh e i suoi compagni fosse equivalente alla decisione di comperare un tubetto di dentifricio, esattamente equivalente a quella di una massimizzazione dell'utilità attesa, ma non è facile dire esattamente che cosa c'è di sbagliato in quell'assunzione, e nel corso di questo libro vorrei provare a farlo. Come prima formulazione intuitiva non possiamo dire che questo. Nella razionalità umana, contrariamente a quella delle scimmie, c'è una distinzione tra le ragioni per agire, che riguardano interamente il soddisfacimento di alcuni desideri piuttosto che altri, e le ragioni che sono indipendenti dai desideri. La distinzione di base tra i diversi tipi di ragioni per l'azione è quella che riguarda le ragioni riferite a ciò che si vuole fare, o ciò che si deve fare per ottenere ciò che si vuole, da un lato, e le ragioni che riguardano ciò che si deve fare indipendentemente da ciò che si vuole, dall'altro.

LE SEI ASSUNZIONI CHE CARATTERIZZANO IL MODELLO CLASSICO

In questo capitolo sosterrò e discuterò sei assunzioni che caratterizzano quello che ho definito il “Modello Classico di razionalità”. Non voglio suggerire che il modello sia unificato, nel senso che se si accetta una proposizione bisogna accettare anche le altre. Al contrario, alcuni autori accettano alcune parti e ne rifiutano delle altre. Ma vorrei sostenere che questo modello costituisce un intero coerente, e influenza gran parte degli scritti contemporanei, in forma sia implicita che esplicita. Inoltre, il modello sviluppa una concezione della razionalità che è quella all’interno della quale mi sono formato come studente di economia e di filosofia morale a Oxford. Non mi sembrava soddisfacente allora, né mi sembra soddisfacente oggi.

1. *Le azioni, se razionali, sono causate da credenze e desideri.*

Le credenze e i desideri hanno la funzione di essere cause e ragioni delle nostre azioni, e la razionalità è in buona parte una questione che riguarda la capacità di coordinare le credenze e i desideri in modo tale che esse causino le azioni “nel modo corretto”.

È importante enfatizzare che qui il termine “causa” indica il significato comune o aristotelico di “causa efficiente”, secondo cui la causa di un evento è ciò che fa sì che esso accada. Queste cause, in un contesto particolare, sono le condizioni sufficienti affinché un evento accada. Affermare che delle specifiche credenze e desideri hanno causato una particolare azione equivale a dire che il terremoto ha causato il crollo dell’edificio.

2. *La razionalità è una questione di osservazione delle regole, le regole speciali che costituiscono la distinzione tra pensiero e comportamento razionale e irrazionale.*

Il nostro compito, in quanto filosofi, è quello di provare a rendere esplicite le regole della razionalità non esplicite che, fortunatamente, la maggior parte delle persone razionali è in grado di seguire inconsciamente. Così come possono parlare inglese senza conoscere le regole della grammatica, o parlare in prosa senza sapere di parlare in prosa, come nel famoso esempio di Monsieur Jourdain, analogamente essi possono comportarsi razionalmente senza conoscere le regole che determinano la razionalità e senza nemmeno essere consapevoli di stare seguendo quelle regole. Ma noi, in quanto teorici, abbiamo come obiettivo lo scoprire e il formulare tali regole.

3. *La razionalità è una facoltà cognitiva separata*

Secondo Aristotele, e secondo l'illustre tradizione filosofica cui egli ha dato inizio, il possesso della razionalità è il tratto che ci caratterizza come esseri umani: l'essere umano è un animale razionale. Oggi il termine di moda per questa facoltà è “modulo”, ma l'idea generale è che gli esseri umani hanno diverse capacità cognitive speciali, una per la visione, una per il linguaggio, ecc. e la razionalità è una di queste facoltà speciali, forse perfino la più caratterizzante tra le nostre capacità umane. Un libro pubblicato di recente propone una riflessione sui vantaggi evolutivi forniti dal possesso di questa facoltà.³

4. *Casi apparenti di debolezza della volontà, ciò che i Greci chiamavano akrasia, possono sorgere solo nei casi in cui c'è qualcosa di sbagliato negli antecedenti psicologici dell'azione*

Poiché le azioni razionali sono causate da credenze e desideri, e le credenze e i desideri normalmente causano l'azione portando dapprima alla formazione di un'intenzione, i casi apparenti di debolezza della volontà richiedono una spiegazione speciale. Com'è possibile che un agente possa avere delle credenze e dei desideri appropriati, elaborare il tipo di intenzione adeguato, e tuttavia non compiere l'azione stessa? La spiegazione tradizionale è che i casi apparenti di *akrasia* comprendono tutti i casi in cui l'agente non aveva, di fatto, il tipo appropriato di antecedente all'azione. Poiché le credenze e i desideri e, in secondo luogo, le intenzioni, sono delle cause, se esse vengono messe assieme razionalmente, l'azione si verificherà per necessità causale. Così, nei casi in cui l'azione non si realizza, deve esserci stato qualcosa di sbagliato nelle cause.

La debolezza della volontà ha sempre costituito un problema per il Modello Classico, e su di essa si è sviluppata un'ampia letteratura,⁴ ma la debolezza della volontà è sempre stata considerata come qualcosa di molto strano e difficile da spiegare, qualcosa che può accadere solamente in circostanze strane o bizzarre. La mia prospettiva personale, che spiegherò più avanti, è che l'*akrasia* negli esseri razionali è comune quanto lo è il vino in Francia. Chiunque abbia mai provato a smettere di fumare, a perdere peso, o a bere di meno alle feste sa di che cosa sto parlando.

³ Robert Nozick, *The Nature of Rationality*, Princeton: Princeton University Press, 1993, tr. it. Robert Nozick, *La natura della razionalità*, tr. it. Feltrinelli, Milano 1995.

⁴ Per un'antologia su questo soggetto, si veda *Weakness of Will*, edito da G.W.Mortimore, London: Macmillan St.Martin's Press, 1971.

5. *La ragione pratica deve iniziare con un inventario degli scopi primari dell'agente, incluse le mete e i desideri fondamentali, gli obiettivi e i propositi; e questi non sono essi stessi soggetti a costrizioni razionali*

Per potersi impegnare nell'attività del ragionamento pratico, un agente deve, per prima cosa, disporre di un insieme di cose che lui o lei desiderano, o dei valori; successivamente, il ragionamento pratico si sviluppa come problema che riguarda l'individuazione del modo migliore per soddisfare questo insieme di desideri e di valori. Possiamo fissare questo punto affermando che affinché il ragionamento pratico disponga di un qualche campo entro cui operare, l'agente deve disporre di un insieme di desideri primari, dove i desideri sono intesi in senso in generale, in modo che le valutazioni dell'agente, siano esse morali, estetiche o di altro tipo, valgano come desideri. Ma in mancanza di un tale insieme di desideri da cui prendere avvio, non vi è alcuno scopo per la ragione, poiché la ragione riguarda la capacità di prefigurare che cos'altro si debba desiderare, dal momento che già si desidera qualcosa. E così i desideri primari non sono essi stessi soggetti a costrizioni razionali.

Il modello della ragione pratica è qualcosa che funziona nel modo seguente: supponete di voler andare a Parigi, e pensate al modo migliore di farlo. Potreste andare in nave, o in kayak, o prendere un aereo e, alla fine, dopo aver esercitato la vostra ragione pratica, decidete di prendere l'aereo. Ma se questo è il solo modo in cui la ragione pratica può operare, prefigurando, cioè, i “mezzi” e i “fini”, ne seguono due cose: per prima cosa, non può esservi alcuna ragione per l'azione che non nasca dai desideri, in senso generale. Cioè, non possono esserci ragioni per l'azione indipendenti dai desideri (*desire-independent*). In secondo luogo, quei desideri iniziali o primari non possono essi stessi essere valutati razionalmente. La ragione riguarda sempre i mezzi, mai i fini.

Questa tesi, secondo la quale non possono esserci ragioni per l'azione indipendenti dai desideri, costituisce il nucleo centrale del Modello Classico. Si è soliti interpretare la frase di Hume secondo cui “la ragione è e dovrebbe essere la schiava delle passioni” come asserente tale tesi; e la stessa tesi è supportata da molti autori recenti. Per esempio, Herbert Simon scrive che “la ragione è un fatto puramente strumentale: non può dirci *dove* andare, al massimo può suggerirci *come* fare a raggiungerlo. La ragione è un'arma a disposizione di

chiunque e che può essere impiegata per qualsiasi scopo, buono o cattivo che sia”.⁵ Bertrand Russell è ancora più sintetico: “la ragione dispone di un significato perfettamente chiaro e conciso. Essa indica la scelta del giusto mezzo in vista di uno scopo che si desidera raggiungere. Non ha niente a che fare con la scelta di questi scopi.”⁶

6. *L'intero sistema della razionalità funziona solo se l'insieme dei desideri primari è consistente*

Una tipica espressione di questa visione è fornita da Jon Elster: “le credenze e i desideri possono difficilmente essere delle ragioni per l'azione se esse non sono consistenti. Esse non devono implicare contraddizioni logiche, concettuali o pragmatiche”.⁷ È facile osservare perché questo risulta plausibile: se la razionalità riguarda il ragionamento logico, non può esservi alcuna inconsistenza o contraddizione all'interno degli assiomi. Una contraddizione implica qualsiasi cosa, quindi, se avessimo una contraddizione nell'insieme iniziale dei desideri, ne potrebbe seguire qualsiasi cosa, o almeno così sembra.

ALCUNI DUBBI SUL MODELLO CLASSICO

Potrei proseguire a lungo in questo elenco, e nel corso di questo libro avremo occasione di arricchire la caratterizzazione del Modello Classico. Ma già questo breve elenco ci dà il senso generale di ciò che vorrei esprimere: aprirò dunque la mia discussione fornendo alcune ragioni sul perché penso che ognuna di queste affermazioni sia falsa. Esse possono forse descrivere dei casi particolari, ma non forniscono una teoria generale del ruolo della razionalità nel pensiero e nell'azione.

1. *Le azioni razionali non sono causate da credenze e desideri. In generale solo le azioni irrazionali e non razionali sono causate da credenze e desideri.*

Prendiamo avvio, come premessa iniziale, dall'idea che le azioni razionali siano quelle causate da credenze e desideri. È importante sottolineare che il senso di “causa” è quello ordinario di “causa efficiente”, come nel caso di un'esplosione che ha causato il

⁵ *Reason in Human Affaire*, Stanford, CA: Stanford University Press, 1983, pp. 7-8; *La ragione nelle vicende umane*, tr. it. Il Mulino, Bologna, 1984, p. 38.

⁶ *Human Society in Ethics and Politics*, London: Allen and Unwin, 1954, p. viii, tr. it. Bertrand Russell, *Un'etica per la politica*, Laterza, Roma-Bari 1986, pag. 4.

⁷ *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge: Cambridge University Press, 1983, p. 4.

crollo dell'edificio o di un terremoto che ha causato la distruzione dell'autostrada. Voglio dire che i casi di azioni per le quali le credenze e i desideri antecedenti sono davvero causalmente sufficienti, lontani dall'essere dei modelli di razionalità, sono, in effetti, casi bizzarri e tipicamente irrazionali. Ci sono casi in cui, per esempio, l'agente è succube di un'ossessione o di un vizio e non può fare altro che agire in base al suo desiderio. Ma in un caso tipico di decisione razionale, quando, per esempio, provo a decidere per quale candidato votare, io ho la possibilità di scegliere e prendo in considerazione ragioni differenti per operare una scelta tra le alternative a mia disposizione. Posso tuttavia impegnarmi in questa attività solamente se assumo che il mio insieme di credenze e desideri non sia di per sé causalmente sufficiente a determinare la mia azione. L'intervento della razionalità presuppone che vi sia un divario (*gap*)⁸ tra l'insieme di stati intenzionali sulla base dei quali prendo la mia decisione, e la reale effettuazione della decisione. Ovvero, se non presuppongo che vi sia un divario, non può avere inizio il processo della decisione razionale. Per capire questo punto bisogna considerare i casi in cui questo salto non c'è, casi in cui la credenza e il desiderio sono davvero causalmente sufficienti. È il caso, per esempio, in cui il tossicodipendente ha un bisogno irresistibile di assumere eroina e crede che ciò che ha davanti sia eroina; e quindi la assume in modo costrittivo. In un tale caso la credenza e il desiderio sono sufficienti a determinare l'azione, poiché il tossicodipendente non può controllarsi. Ma questo difficilmente potrebbe costituire un modello di razionalità. Tali casi sembrano piuttosto essere al di fuori dell'ambito generale della razionalità.

Nel caso normale di un'azione razionale, dobbiamo presupporre che l'insieme di credenze e desideri antecedente non sia causalmente sufficiente a determinare l'azione. Questo è un presupposto del processo di deliberazione ed è assolutamente indispensabile per l'applicazione della razionalità. Noi presupponiamo che vi sia un divario tra le "cause" dell'azione, le credenze e i desideri, e l'"effetto", ovvero l'azione. Questo divario ha un nome tradizionale. E' chiamato "libero arbitrio". Per impegnarci in una decisione razionale dobbiamo presupporre l'esistenza del libero arbitrio. Infatti, come vedremo più avanti, dobbiamo presupporre la presenza del libero arbitrio in qualsiasi attività razionale. Non possiamo evitare tale presupposizione, poiché persino il rifiuto ad impegnarsi in una decisione razionale ci risulta intelligibile come rifiuto soltanto se lo consideriamo come un esercizio della libertà. Per chiarire questo punto, considerate i seguenti esempi. Immaginate

⁸ *N.d.t.* Tradurremo il termine inglese "*gap*" con l'italiano "divario". L'estensione semantica del termine, sia in inglese che in italiano, è tuttavia più ampia e indica anche "differenza", "salto", "spazio vuoto".

di entrare in un ristorante, e che il cameriere vi porti il menu. Avete a disposizione una scelta, diciamo, tra una braciola di vitello e un piatto di spaghetti; voi non potete dire: “Guardi, io sono un determinista: ‘che sarà, sarà’. Semplicemente aspetterò e vedrò che cosa ordino! Aspetterò di vedere che cosa le mie credenze e i miei desideri causeranno”. Tale rifiuto ad esercitare la vostra libertà vi appare comprensibile solamente in quanto esercizio della libertà. Kant lo mise in rilievo molto tempo fa: nell’azione volontaria non c’è alcuna possibilità di pensare al di fuori della nostra libertà perché il processo stesso della deliberazione si fonda sul presupposto della libertà, sul presupposto cioè che vi sia un divario tra le cause, nella forma di credenze, desideri e altre ragioni, e la decisione effettiva che si compie.

Se volessimo parlare di questo in forma esatta, credo dovremmo dire che vi sono (almeno) tre divari. Primo: c’è un divario, o una differenza, nel prendere delle decisioni razionali, quando si cerca di prendere una decisione riguardo a ciò che si intende fare. Qui il divario è tra le ragioni che portano alla decisione e la decisione che effettivamente si prende. Secondo: c’è un divario tra la decisione e l’azione. Proprio come le ragioni per la decisione non erano causalmente sufficienti per produrre la decisione, così la decisione non è causalmente sufficiente per produrre l’azione. Dopo che si è presa la decisione, arriva il momento in cui bisogna realmente compiere l’azione. E, ancora una volta, non ci si può astenere e lasciare che la decisione causi l’azione, e nemmeno si può lasciare che le ragioni causino la decisione. Per esempio, supponiamo che vi siate decisi a votare per il candidato Jones. Entrate nella cabina elettorale con questa decisione fissata nella vostra mente, ma una volta lì dentro dovete ancora compiere l’azione. E talvolta, a causa del secondo divario, semplicemente non riuscite a farlo. Per diverse ragioni possibili – o forse per nessuna – non fate la cosa che avevate deciso di fare.

C’è un terzo divario che si presenta per le azioni e le attività che si estendono nel tempo, un divario tra l’inizio dell’azione e la sua continuazione fino al completamento. Supponete, per esempio, di aver deciso di imparare il portoghese, o di attraversare a nuoto la Manica, o di scrivere un libro sulla razionalità. C’è, per prima cosa, il divario tra le ragioni che portano alla decisione e la decisione stessa; in secondo luogo c’è il divario tra la decisione e il cominciamento dell’azione; terzo, c’è il divario tra l’inizio del compito e la sua continuazione fino al completamento. Anche dopo aver iniziato l’azione non si possono lasciare operare le cause da sole; si deve compiere uno sforzo volontario continuo per proseguire l’azione o l’attività fino al suo completamento.

A questo punto della discussione voglio sottolineare due punti: l'esistenza del divario (o dei divari) e la sua centralità per l'argomento della "razionalità".

Quali sono gli argomeni a favore dell'esistenza del divario (o dei divari)? Svilupperò questi argomenti in maggiore dettaglio nel capitolo 3; per gli scopi attuali possiamo dire che gli argomenti più semplici sono quelli che ho appena fornito. Considerate qualsiasi situazione che riguardi il prendere decisioni razionali e l'agire: vi troverete di fronte alla sensazione di poter disporre di numerose possibilità alternative e il vostro agire e deliberare acquisterà senso soltanto sulla base del presupposto che esistano tali possibilità alternative. Mettete a confronto queste situazioni con quelle in cui non c'è una tale sensazione di possibilità alternative. Per esempio, nella situazione in cui vi trovaste succubi di una rabbia incontrollabile, al punto di essere, come si suol dire, completamente fuori di senno, non avreste la sensazione di poter agire in modo diverso da quello.

Un altro modo per rendersi conto dell'esistenza del divario è notare che, in una situazione che comporta una decisione, spesso si dispone di molteplici ragioni differenti per compiere una certa azione, tuttavia si agisce in base ad una certa ragione e non ad altre: si conosce senza osservazione (*without observation*) la ragione in base alla quale si è agito. Questo è un fatto rilevante e va notata la curiosa locuzione che usiamo per descriverlo: *si agisce in base a* quella particolare ragione. Supponiamo, per esempio, che aveste a disposizione un considerevole numero di ragioni per votare sia a favore che contro Clinton alle elezioni presidenziali. Pensavate che sarebbe stato un miglior presidente per l'economia, ma peggiore per la politica estera. Vi piaceva il fatto che avesse frequentato il vostro vecchio College ma non vi piaceva il suo stile personale. Alla fine avete votato per lui perché aveva frequentato il vostro vecchio College. Le ragioni non hanno avuto alcun effetto su di voi. O meglio, *avete scelto* una ragione ed avete agito in base ad essa. Avete reso effettiva quella ragione *agendo in base ad essa*.

Questo è il motivo per cui, incidentalmente, la spiegazione della vostra azione e della sua giustificazione potrebbe non essere la stessa. Supponete che vi venga chiesto di motivare il vostro voto per Clinton; potreste farlo appellandovi alla sua grande capacità di gestire l'economia. Ma potrebbe succedere che la ragione reale in base alla quale avete agito fosse data dal fatto che Clinton aveva frequentato il vostro vecchio College a Oxford e voi pensavate che "la fedeltà verso il College viene prima di tutto". E una cosa da sottolineare riguardo a questo fenomeno è che in un caso normale voi sapete senza osservazione quale

ragione era efficace, proprio perché voi stessi l'avete resa efficace. Ciò vuol dire che una ragione per l'azione è efficace solo se siamo noi stessi a renderla tale.

La comprensione del divario è essenziale per l'argomento della razionalità perché la razionalità può operare solamente all'interno di tale divario. Sebbene il concetto di libertà e il concetto di razionalità siano alquanto differenti, l'estensione della razionalità corrisponde esattamente a quella della libertà. L'argomento più semplice a favore di questo punto è che la razionalità è possibile solo dove è possibile l'irrazionalità, e questo requisito implica la possibilità di scegliere tra differenti opzioni razionali tanto quanto tra diverse opzioni irrazionali. L'obiettivo di tale la scelta è costituito dal divario in questione. La tesi che la razionalità possa operare solamente all'interno del divario è vera sia per la ragione teorica che per la ragione pratica, ma la parte che riguarda la ragione teorica è un punto più sottile da analizzare, lo riserverò quindi per dopo e mi concentrerò ora sulla ragione pratica.

Nel corso di questo libro parlerò a lungo del divario e in un certo senso il libro riguarda proprio il divario, perché il problema della razionalità è un problema che ha a che fare con il divario. A questo punto pongo altre due questioni.

Primo: che cosa colma il divario? Niente. Niente colma il divario: voi prendete la decisione di fare qualcosa o cambiate idea e fate ciò che stavate per fare, o portate a termine la decisione che avevate preso in precedenza, o andate avanti, o smettete di portare avanti un certo progetto che avevate intrapreso.

Secondo: anche se apparentemente disponiamo di tutte queste esperienze, tutto questo non potrebbe essere un'illusione? Sì, potrebbe esserlo. Le nostre esperienze legate al divario non si auto-garantiscono. Sulla base di quanto ho affermato finora, il libero arbitrio potrebbe ancora essere una pesante illusione. La realtà *psicologica* del divario non garantisce una corrispondente realtà *neurobiologica*. Esplorerò questo argomento nel capitolo 9.

2. *La razionalità non è una questione che riguarda il seguire regole razionali.*

Passiamo ora alla seconda tesi del Modello Classico, ovvero la tesi secondo la quale la razionalità è una questione che riguarda le regole, per cui noi pensiamo e ci comportiamo razionalmente solo in relazione al fatto che pensiamo e agiamo in accordo con queste regole. Se si chiedesse di giustificare questa tesi, penso che i teorici più tradizionali si appellerebbero semplicemente alle regole della logica. Un tipo di argomento ovvio che un

difensore del Modello Classico potrebbe presentare sarebbe, diciamo, un argomento che presenta un semplice *modus ponens*:

Se stanotte piove, il terreno sarà umido.

Stanotte pioverà.

Quindi il terreno sarà umido.

Ora alla richiesta di giustificare questa inferenza, la tentazione sarebbe quella di appellarci alla regola del *modus ponens*: dato p , e se p allora q , ciò implica q .

$$(p \ \& \ (p \rightarrow q)) \rightarrow q$$

Ma questo è un errore fatale. Quando si sostiene ciò, si finisce nella morsa del paradosso di Lewis Carroll⁹. Ora vi ricorderò come si svolge: Achille e la tartaruga stanno discutendo e Achille dice (questo non è l'esempio originale di Carroll, ma sottolinea il medesimo punto): “se stanotte piove, il terreno sarà umido, pioverà stanotte, quindi il terreno sarà umido”. E la tartaruga dice: “Bene! Mettilo per iscritto, metti tutto per iscritto!”. E una volta che Achille ha fatto questo, dice: “non vedo come riesci ad ottenere da quanto precede il ‘quindi’ tutto quello che segue. Che cosa ti obbliga a fare o semplicemente a giustificare questa mossa?”. Achille dice: “ebbene, questa mossa si basa sulla regola del *modus ponens*, la regola secondo la quale dato p , e se p allora q , tutto ciò implica q .” “Bene!” dice la tartaruga: “Mettilo per iscritto, insieme a tutto il resto!”. E, una volta che Achille l’ha fatto, la tartaruga dice: “Bene! Abbiamo tutto questo per iscritto, ma ancora non capisco come ottieni la conclusione secondo la quale il terreno sarà umido”. “Ebbene, non lo vedi?” dice Achille “ogni volta che si ha p , e se p allora q , allora hai la regola del *modus ponens* che dice che quando si ha p , e se p allora q , si può inferire q , allora si può inferire q .” “Bene!” dice la tartaruga “ora metti tutto per iscritto”. Potete ben capire come prosegue. Abbiamo a che fare con un regresso all’infinito.

Il modo per evitare un regresso all’infinito è rifiutarsi di compiere il primo fatale movimento di supporre che la regola del *modus ponens* giochi *qualsiasi ruolo in qualunque caso* per rendere valida l’inferenza. La derivazione non viene resa valida dalla regola del *modus ponens*; piuttosto, l’inferenza è perfettamente valida in quanto essa regge senza il

bisogno di alcun aiuto esterno. Sarebbe più accurato dire che la regola del *modus ponens* ottiene la sua validità dal fatto che essa esprime il modello di un infinito numero di inferenze che sono valide indipendentemente. L'argomento reale non ottiene la sua validità da una qualsivoglia fonte esterna: se esso è valido, lo è soltanto perché le premesse implicano la conclusione. Poiché i significati delle parole stesse sono sufficienti a garantire la validità dell'inferenza, possiamo formalizzare un modello che descriva un numero infinito di tali inferenze. Ma l'inferenza non deriva la sua validità dal modello. La cosiddetta regola del *modus ponens* è solo l'enunciato di un modello di un numero infinito di tali inferenze che sono valide indipendentemente dal resto. Ricordate: *se pensate di aver bisogno di una regola per inferire q da p e (se p allora q), allora avrete bisogno anche di una regola per inferire p da p.*

Ciò che funziona per questo argomento funziona per qualsiasi argomento deduttivo valido. La validità logica non deriva dalle regole della logica.

È importante comprendere questo punto in modo preciso. Si è soliti affermare che l'errore di Achille è stato quello di trattare il *modus ponens* come un'altra premessa e non come una regola. Ma questo è sbagliato. Anche qualora lo mettesse per iscritto come una regola e non come una premessa, vi sarebbe ancora un regresso all'infinito. È ugualmente sbagliato (e in effetti è il medesimo errore) dire che la derivazione deriva la sua validità sia dalle premesse che dalla regola di inferenza.¹⁰ È corretto affermare che le regole della logica non giocano alcun ruolo per quanto riguarda la validità delle inferenze valide. Gli argomenti, se sono validi, devono essere validi di per sé.

In effetti, di fronte a questo punto siamo resi ciechi dalla nostra raffinatezza esasperata, poiché i successi della teoria della dimostrazione sono stati così rilevanti e hanno avuto tali successi in campi come la scienza informatica, che pensiamo che l'analogo sintattico del *modus ponens* sia realmente la stessa cosa della "regola" della logica. Ma si tratta di due cose alquanto differenti. Supponete di avere una regola reale la quale, ogni volta che vedete qualcosa, o che il vostro computer "vede", vi presenti un simbolo con questa forma:

p

seguito da un simbolo con questa forma:

⁹ L. Carroll, "What Achilles Said to the Tortoise", *Mind* 4: 278-280, aprile 1985.

¹⁰ Per un esempio di questa posizione si veda Peter Railton, "On the Hypothetical and the Non-Hypothetical in Reasoning about Belief and Action", pp. 53-79 in G. Cullity and B. Gaut, *Ethics and Practical Reason*, Oxford, Oxford University Press, 1997, spec. pp. 76-79.

$p \rightarrow q$,

e voi, o il computer, ne scrivete uno con questa forma:

q ,

avreste una regola reale che potete seguire e che potete programmare nella macchina in modo da provocare causalmente le sue operazioni. Questa è una dimostrazione teorica analoga alla regola del *modus ponens*, ed essa è realmente indipendente, poiché mostra che la regola che compare sopra non è altro che un insieme di simboli privi di significato. Ovvero, che tale regola non si riferisce a elementi formali.

Non siamo dunque capaci di notare il fatto che nei ragionamenti che facciamo nella vita reale la regola del *modus ponens* non gioca alcun ruolo giustificativo. Possiamo produrre modelli di dimostrazione teorica o sintattici, in cui il modello rispecchia esattamente i processi essenziali o contenutistici dell'effettivo ragionare umano. E certamente, come tutti sappiamo, con i modelli si può fare molto. Se si dispone di una sintassi corretta, allora si può attivare una semantica sin all'inizio, ed essa andrà avanti liberamente e, alla fine del processo, otterremo una semantica corretta poiché abbiamo operato le corrette trasformazioni sintattiche.

Esistono alcuni problemi famosi, tra i quali il più famoso è il teorema di Gödel, ma se li lasciamo da parte, vedremo che la raffinatezza delle nostre simulazioni nei modelli meccanici di ragionamento ci farà dimenticare il contenuto semantico. Ma nel ragionamento della vita reale è proprio il contenuto semantico ciò che garantisce la validità dell'inferenza, *non le regole sintattiche*.

Riguardo al paradosso di Lewis Carroll vi sono due considerazioni filosofiche importanti da mettere in luce. La prima, con la quale vi ho annoiato, è che la regola non gioca alcun ruolo in relazione alla validità dell'inferenza. La seconda riguarda il divario. *Abbiamo bisogno di distinguere tra l'implicazione e la validità come relazioni logiche da una parte, e l'inferenza come attività umana, dall'altra*. Nel caso che abbiamo preso in considerazione, le premesse implicano la conclusione, e così l'inferenza è valida. Ma non vi è nulla che obblighi un qualsiasi essere umano reale a compiere quell'inferenza. Si ha lo stesso divario per l'attività umana del compiere inferenze come per qualsiasi altra attività volontaria. Persino se convincessimo sia Achille che la tartaruga che l'inferenza era valida di per sé e che la regola del *modus ponens* non conferisce nessuna validità all'inferenza, in ogni caso, la tartaruga potrebbe rifiutare ancora, irrazionalmente, di compiere l'inferenza. Il divario si applica persino alle inferenze logiche.

Non sto dicendo che non ci dovrebbero essere delle regole che ci aiutano a prendere decisioni razionali. Al contrario, ci sono molte famose regole del genere e perfino delle massime. Eccone alcune: “un punto in tempo ne salva cento”, “pensaci prima di fare un passo”, “ride bene chi ride ultimo”. E il mio preferito: “*le cœur a ses raisons que la raison ne connaît pas*”. Ciò che sto dicendo è che la razionalità non è costituita come un insieme di regole, e la razionalità del pensiero, così come dell’azione, non è definita da un insieme di regole. La struttura degli stati intenzionali e le regole costitutive degli atti linguistici già includono le costrizioni della razionalità.

3. *Non esiste una facoltà separata della razionalità.*

In quanto ho detto dovrebbe essere implicito che la razionalità non può essere una facoltà separata, distinta da capacità come il linguaggio, il pensiero, la percezione e le diverse forme dell'intenzionalità, poiché le costrizioni razionali sono già insite, interne, alla struttura dell'intenzionalità in generale e del linguaggio in particolare. Una volta che si hanno degli stati intenzionali, delle credenze e dei desideri, speranze e paure, e, specialmente, una volta che si dispone del linguaggio, allora si hanno già le costrizioni della razionalità. Cioè, se un animale dispone della capacità di formare credenze sulla base delle sue percezioni ed ha la capacità di formare desideri in aggiunta alle credenze, ed ha pure la capacità di esprimere tutto questo in un linguaggio, allora esso già dispone delle costrizioni della razionalità costruite in queste strutture. Per chiarire questo punto con un esempio: non è possibile esprimere alcun enunciato senza impegnarsi in domande come “è vero o è falso?”, “è consistente o inconsistente con le altre cose che ho detto?”. Così, le costrizioni della razionalità non sono una facoltà "extra" che si aggiunge all'intenzionalità e al linguaggio. Una volta che abbiamo l'intenzionalità e il linguaggio, abbiamo già dei fenomeni che possiedono le costrizioni della razionalità, internamente e costitutivamente.

Mi piace pensarla in questo modo: le costrizioni della razionalità dovrebbero essere pensate in forma avverbiale. Esse riguardano il modo in cui noi coordiniamo la nostra intenzionalità. Riguardano il modo in cui coordiniamo le relazioni tra le nostre credenze, i nostri desideri, le nostre speranze, le nostre paure, le percezioni e altri fenomeni intenzionali.

Questa azione di coordinazione presuppone l'esistenza del divario. Presuppone che i fenomeni, in ogni determinato punto, non siano causalmente sufficienti per stabilire la soluzione razionale di un problema. E penso che ora possiamo comprendere perché questo

funziona sia per la ragione teorica che per la ragione pratica. Se sollevo la mano e la pongo dinanzi al mio viso, non vi è alcun divario nella visione della mia mano, perché non posso impedire a me stesso di vedere la mano posta dinanzi al mio viso, se la luce è sufficiente e la mia vista è buona. Non sta a me decidere. Non si pone dunque la questione se una tale percezione sia razionale o irrazionale. Ma supponiamo che io mi rifiuti di accettare questo fatto: “tu mi dici che là c’è una mano, ma io mi rifiuto fermamente di accettare questa affermazione”. In questo caso, sorge la questione della razionalità e potremmo sostenere che in questa situazione io mi comporto in modo irrazionale.

Vorrei ora sottolineare un punto che avevo affrontato in precedenza. Si può avere la razionalità soltanto dove si ha la possibilità dell’irrazionalità. E con percezioni semplici e dirette non si hanno né razionalità né irrazionalità. Queste entrano in gioco soltanto dove vi è un divario, dove l’esistenza dei fenomeni intenzionali non è di per sé sufficiente a causarne il risultato, e questi sono i casi in cui si deve decidere ciò che si sta per fare o per pensare.

Questo spiega perché nei casi in cui il comportamento delle persone è determinato da condizioni causali sufficienti non abbiamo alcun obiettivo nella valutazione razionale. Per esempio, non molto tempo fa mi trovavo ad una riunione di commissione, e una persona della quale avevo profondo rispetto votò nel modo più stupido possibile. In seguito gli chiesi: “come puoi aver votato in quel modo riguardo a quella questione?” Ed egli disse: “Beh, sono proprio inguaribilmente *politicamente corretto*. Non posso impedirmi di esserlo”. La sua affermazione equivaleva a dire che la sua decisione, in questo caso, era al di fuori degli obiettivi di una valutazione razionale, poiché l’apparente irrazionalità risultava dal fatto che egli non aveva nessuna scelta, che le cause erano causalmente sufficienti.

4. *La debolezza della volontà è una forma comune, naturale, di irrazionalità. È una conseguenza naturale del divario.*

Per il Modello Classico, i casi di debolezza della volontà sono, in senso stretto, impossibili. Se gli antecedenti dell’azione sono razionali e causali e le cause costituiscono un insieme di condizioni sufficienti, allora l’azione deve necessariamente verificarsi. Ne segue che se non si realizzano le cose che si era stabilito di realizzare, questo può essere dovuto solamente al fatto che c’era qualcosa di sbagliato nel modo in cui abbiamo elaborato

gli antecedenti dell'azione. La vostra intenzione non era il corretto tipo di intenzione,¹¹ oppure non eravate completamente impegnati, in senso morale, verso azioni rispetto alle quali sostenevate di esservi impegnati.¹²

Voglio invece affermare, al contrario, che per quanto perfettamente si strutturino gli antecedenti dell'azione, la debolezza della volontà è sempre possibile. Ecco in che modo: in ogni momento della nostra vita da svegli, siamo messi di fronte a un'infinita gamma di diverse possibilità. Posso sollevare il mio braccio destro o posso sollevare quello sinistro; posso mettere il cappello sulla mia testa o posso sventolarlo. Posso bere dell'acqua o non berla. In forma più radicale, posso uscire dalla camera e andare a Timbuktù, o ritirarmi in un monastero, o fare moltissime altre cose. Ho a disposizione diverse possibilità aperte a qualsiasi direzione. Ora, sicuramente nella vita reale vi saranno delle restrizioni costituite dal mio *Sfondo* (*Background*), dai miei limiti biologici e dalla cultura all'interno della quale sono cresciuto. Lo *Sfondo* restringe la mia percezione delle possibilità che mi si aprono dinanzi in qualsiasi momento. Io non posso, per esempio, nella vita reale, immaginare di fare ciò che fece San Simeone lo Stilite. Egli trascorse trentacinque anni in cima a una colonna, seduto su di una piccola piattaforma, per la gloria di Dio. Questa non è un'opzione che posso prendere seriamente in considerazione. Ma ho a disposizione una gamma infinita di opzioni reali che sono in grado di percepire come opzioni. La debolezza della volontà sorge semplicemente dal fatto che in ogni punto il divario mi fornisce un'ampia gamma indefinita di scelte, e tra queste alcune mi sembreranno attraenti anche dopo aver già preso la decisione di escluderle. Qualunque sia il modo in cui si strutturano le cause dell'azione, nella forma di stati intenzionali antecedenti – credenze, desideri, scelte, decisioni, intenzioni –, nel caso delle azioni volontarie le cause non costituiscono ancora delle condizioni sufficienti, e questo apre la strada alla debolezza della volontà.

Ritengo sia un tratto negativo della nostra tradizione filosofica fare apparire la debolezza della volontà come qualche cosa di strano o di estremamente bizzarro; penso che essa sia invece un fenomeno molto comune nella nostra vita reale. Dedicherò il capitolo 7 a questo problema, quindi ora non mi soffermerò ulteriormente su di esso.

¹¹ Donald Davidson, "How Is Weakness of the Will Possible?" *Essays on Actions and Events*, Oxford, Clarendon Press, Oxford University Press, New York, 1980, "Com'è possibile la debolezza della volontà?" in *Azioni ed eventi*, tr. it. Il Mulino, Bologna, 1992, 2000.

¹² R.M. Hare, *The Language of Morals*, Oxford, Oxford University Press, 1952, tr. it. R.M. Hare, *Il linguaggio della morale*, tr. it. Ubaldini, Roma 1968.

5. *Contrariamente al Modello Classico, vi sono ragioni indipendenti dai desideri per l'azione.*

La quinta tesi del Modello Classico che voglio mettere in discussione ha una lunga storia all'interno della nostra tradizione filosofica. L'idea è la seguente: un'azione razionale può essere motivata soltanto da un desiderio, dove "desiderio" è inteso in senso ampio, fino ad includere valori morali condivisi e tipi diversi di valutazioni. Non tutti i desideri devono essere riferiti al proprio io, ma per qualsiasi processo razionale di deliberazione deve esserci qualche desiderio che l'agente aveva precedentemente al processo, altrimenti non vi sarebbe nulla da cui iniziare a ragionare. Se non si avesse un insieme di desideri precedenti, non vi sarebbe nessuna base sulla quale sviluppare il proprio ragionamento. Così, non può esserci alcun ragionamento riguardo ai fini, ma solo riguardo ai mezzi. Una versione sofisticata di questa visione si trova nell'opera di Bernard Williams,¹³ il quale sostiene che non può esserci alcuna ragione "esterna" perché un agente agisca. Qualsiasi ragione che costituisce una ragione per l'agente deve appellarsi a qualcosa di "interno", al suo "insieme motivazionale". Questo, secondo la mia terminologia, equivale a dire che non può esserci nessuna ragione per l'azione indipendente dai desideri.

Criticherò questa visione in forma più ampia e in maggiore dettaglio più avanti, per il momento vorrei soffermarmi su di una critica in particolare. Questa prospettiva porta alla seguente conseguenza assurda: in ogni particolare momento della vita di una persona, non importa quali siano i fatti e non importa che cosa uno ha fatto in passato o che cosa conosca del futuro, nessuno può avere una ragione per fare qualcosa fino a quando, proprio in quel momento, non emerge un elemento dell'insieme motivazionale di quella persona, un desiderio in senso ampio, che spinga a fare quella particolare cosa, o un desiderio per cui il fare quella cosa sarebbe un "mezzo" per realizzare quel "fine", cioè, un mezzo per soddisfare quel desiderio.

Perché ciò è assurdo? Ebbene, provate ad applicarlo in esempi della vita reale. Supponete di entrare in un bar e di ordinare una birra. Il cameriere vi porta la birra e voi la bevete. Successivamente il cameriere vi porta il conto e voi gli dite: "ho considerato il mio

¹³ "External and Internal Reasons", ristampato nel suo *Moral Luck: Philosophical Papers 1973-1980* Cambridge, Cambridge University Press, 1981, pp. 101-113, tr. it. "Ragioni interne ed esterne" in *Sorte morale*, Il Saggiatore, Milano, 1987, pp. 133-147. Williams nega che il suo modello sia riferito soltanto al ragionamento basato sull'opposizione fini-mezzi, ma gli altri tipi di casi che considera, come l'inventare corsi alternativi di azioni, non mi sembrano alterare la struttura di base del suo modello, basato sull'opposizione fini-mezzi. Si veda il suo "Internal Reasons and the Obscurity of Blame" ristampato nel suo *Making Sense of Humanity and Other Philosophical Papers*, Cambridge University Press, 1998, pp. 38-45.

insieme motivazionale e non ho trovato nessuna ragione interna per pagare questa birra. Nemmeno una. Ordinare e bere la birra è una cosa, trovare qualcosa nel mio insieme motivazionale, è un'altra cosa. I due eventi sono logicamente indipendenti. Pagare la birra non è qualcosa che desidero di per sé, né è un mezzo per un fine, né è costitutivo di un qualche fine che è rappresentato nel mio insieme motivazionale. Ho letto i lavori del Professor Williams, ho anche letto Hume riguardo a questo argomento, ho osservato attentamente il mio insieme motivazionale, ma non riesco a trovarvi nessun desiderio di pagare questo conto! Proprio non ci riesco! E, perciò, secondo tutte le spiegazioni tradizionali della razionalità, non ho nessuna ragione per pagare questa birra. Non è che non ho una ragione abbastanza forte, o che ho delle ragioni in conflitto tra loro: ho zero ragioni. Ho osservato il mio insieme motivazionale, ho fatto un intero inventario, non ho trovato nessun desiderio che mi porterebbe, tramite una scelta chiaramente deliberata, all'azione il pagare la birra.”

Troviamo questo discorso assurdo perché capiamo che quando si ordina una birra e la si beve, se si è una persona sana e razionale, si *crea* intenzionalmente una ragione indipendente dal desiderio, una ragione per fare qualcosa che è noncurante di ciò che si trovava nel nostro insieme motivazionale quando è arrivato il momento di agire. L'assurdità sta nel fatto che, secondo il Modello Classico, l'esistenza di una ragione per cui un agente agisce dipende dall'esistenza di un certo tipo di elemento psicologico nel suo insieme motivazionale; essa dipende dall'esistenza di un desiderio, inteso in senso ampio, proprio in quel momento. E in assenza di quel desiderio, l'agente non ha alcuna ragione, indipendente da tutti gli altri fatti che riguardano lui e la sua storia, e indipendente da quanto egli conosce. Ma nella vita reale la pura conoscenza dei fatti esterni del mondo, come il fatto che avevate ordinato una birra e che l'avevate bevuta, può costituire una ragione schiacciante perché voi paghiate la birra.

La domanda, com'è possibile che vi siano ragioni per l'azione indipendenti dai desideri, è una domanda interessante e non marginale. Io credo che la maggior parte delle spiegazioni tradizionali siano errate. Intendo esporre una discussione estesa di questo punto più avanti nel libro, al capitolo 6, e dunque qui non la discuterò ulteriormente.

Nel Modello Classico vi sono due linee essenziali riguardo a questo particolare aspetto. Per prima cosa, si suppone che il ragionare sia una questione che riguarda i mezzi e non i fini, che non vi sono ragioni esterne per l'azione. Seconda cosa, si sostiene che siamo inclini a credere ad un corollario, secondo il quale i fini primari dell'insieme motivazionale

sono esterni all'ambito della ragione. Ricordiamoci che Hume diceva anche: “non è contrario ai dettami della ragione il preferire la distruzione del mondo intero allo scorticarsi del mio dito mignolo.” Il modo per prendere in considerazione affermazioni del genere è sempre quello di trasporle in un caso della vita reale. Supponete che il presidente degli Stati Uniti dica alla televisione: “mi sono consultato con il Gabinetto e con i capi del Congresso, ed ho deciso che non c’è una ragione per cui io dovrei preferire lo scorticarsi del mio dito mignolo alla distruzione del mondo intero”. Se lo avesse fatto nella vita reale noi avremmo pensato, per usare la terminologia dell’epoca di Hume, che egli aveva “perduto la ragione”. C’è qualcosa di ambiguo nella tesi di Hume e nella tesi generale secondo la quale i fini fondamentali di una persona possono essere qualsiasi cosa, che essi sono totalmente esterni all'ambito della razionalità, e che quando ci si riferisce ai desideri primari, qualunque cosa ha una condizione equivalente ed è ugualmente arbitraria. Non penso che questo sia il modo corretto per affrontare questi problemi.

La tesi secondo la quale non ci sono ragioni per l’azione indipendenti dai desideri, che non ci sono ragioni esterne, è logicamente connessa alla dottrina di Hume secondo la quale non si può derivare un “dovrebbe” da un “è”. Questa è la connessione. Gli enunciati in cui compare il “dovrebbe” esprimono le ragioni per l’azione. Dire che qualcuno dovrebbe fare qualcosa implica che vi sia una ragione per lui per agire in quel modo. Così, la tesi di Hume equivale alla tesi per cui gli enunciati che asseriscono l’esistenza di ragioni per l’azione non possono essere derivati da enunciati che riguardano lo stato reale delle cose. Ma lo stato delle cose è una questione che riguarda come le cose sono effettivamente nel mondo: esso esiste indipendentemente dall’insieme motivazionale dell’agente. Così, sulla base di questa interpretazione, la tesi secondo la quale lo stato di cose del mondo non può implicare l’esistenza di una qualsivoglia ragione nell’insieme motivazionale di un agente (non si può derivare “dovrebbe” da “è”) è strettamente collegata alla tesi che non ci sono fatti nel mondo, indipendenti dall’agente, che costituiscano di per sé le ragioni per l’azione (non vi sono ragioni esterne). Hume sostiene, in effetti, che non possiamo ottenere dei valori da dei fatti; Williams sostiene che non possiamo ottenere delle motivazioni dai semplici fatti esterni di per sé. Il punto di connessione è dato dal fatto che accettare un valore significa accettare una motivazione. Comunque noi interpretiamo le due posizioni, penso sia possibile dimostrare che esse sono entrambe false, e intendo discutere questo problema in dettaglio nel corso di questo libro.

6. *Le ragioni per l'azione inconsistenti sono comuni e, in effetti, inevitabili. Non esiste un requisito razionale per il quale la decisione razionale deve partire da un insieme consistente di desideri o di altre ragioni primarie per l'azione.*

L'ultimo punto che voglio trattare è la questione della consistenza, o coerenza. Così come per l'argomento della debolezza della volontà, questa parte del Modello Classico – la tesi secondo cui l'insieme dei desideri primari, da cui prende avvio il ragionamento, deve essere coerente – non mi sembra soltanto un po' falsa, bensì radicalmente errata. Mi sembra che una caratteristica essenziale del ragionamento pratico sia quella di giudicare tra desideri conflittuali, desideri inconsistenti e altri tipi di ragioni. Proprio oggi, in questo momento, vorrei tantissimo essere a Parigi, ma desidero moltissimo anche rimanere a Berkeley. E questa non è una situazione bizzarra; piuttosto, mi sembra del tutto normale che noi disponiamo di un insieme inconsistente di fini. Data una premessa ulteriore secondo la quale io non posso essere contemporaneamente a Berkeley e a Parigi, io mi ritrovo ad avere un insieme incoerente di desideri; e il compito della razionalità, il compito della ragione pratica, è quello di cercare di trovare un qualche modo per giudicare tra questi diversi fini inconsistenti. Normalmente nel ragionamento pratico bisogna capire come è possibile rinunciare al soddisfacimento di alcuni desideri per soddisfarne altri. Nella letteratura filosofica tradizionale, la via d'uscita da questo problema è dire che la razionalità non riguarda i desideri, bensì le *preferenze*. La deliberazione razionale deve iniziare con una programmazione bene ordinata delle preferenze. Il problema di questa risposta è che nella vita reale la deliberazione riguarda in gran parte proprio il formare un insieme di preferenze. E un insieme di preferenze ben ordinate è, solitamente, il *risultato* di una deliberazione ben riuscita, non si tratta della sua *condizione di partenza*. Che cosa preferisco, essere a Berkeley o a Parigi? Bene, dovrei rifletterci.

E persino dopo che si è presa una decisione e si sceglie “Bene, vado a Parigi”, questa stessa decisione introduce tutta una serie di altri conflitti. Volete andare a Parigi, ma non volete fare la coda negli aeroporti, non volete mangiare il cibo che vi danno in aereo, non volete sedervi vicino a qualcuno che cerca di mettere il suo ginocchio proprio dove voi volete mettere il vostro, e così via. Ci sono tutti quei generi di cose che non vorreste che accadessero, ma che sapete che accadranno una volta che cercherete di realizzare concretamente la vostra decisione di andare a Parigi, e di andarci in aereo. Nel soddisfare un desiderio, ne frustrate degli altri. Il punto che voglio sottolineare è che c'è una lunga tradizione associata al Modello Classico secondo la quale delle ragioni inconsistenti per

l'azione, così come degli obblighi inconsistenti, sono considerate filosoficamente strane o inusuali. Spesso, nell'ambito di questa tradizione, si cerca di risolvere le incoerenze dicendo che alcuni degli obblighi apparentemente inconsistenti non sono veri e propri obblighi, ma semplici obblighi *prima facie*. Ma il prendere decisioni razionali riguarda, innanzitutto, la capacità di scegliere tra ragioni per l'azione conflittuali, e si ha un reale conflitto di obblighi solo dove ci sono degli obblighi genuini. Esiste una questione seria ed è quella di capire come possano esserci ragioni per l'azione logicamente inconsistenti ma ugualmente valide, e perché la ragione pratica debba implicare dei conflitti tra queste ragioni valide, ma logicamente inconsistenti. Ritorrò su questo problema, in maggiore dettaglio, nei capitoli seguenti.

Lo scopo di questo capitolo era quello di introdurre l'argomento del libro mettendo in luce alcuni dei principi costitutivi della tradizione che vorrei superare, ed esprimendo, in forma preliminare, alcune delle mie obiezioni a tale tradizione. Poiché abbiamo iniziato il capitolo con le scimmie di Köhler, concluderemo con esse. Secondo il Modello Classico la razionalità umana è un'estensione della razionalità dello scimpanzé. Noi siamo degli scimpanzé parlanti, estremamente intelligenti. Ma io penso che vi siano delle differenze fondamentali tra la razionalità umana e la ragione strumentale degli scimpanzé. La più grande differenza tra gli esseri umani e il resto del regno animale, per quanto riguarda la razionalità, è la nostra capacità di creare, di riconoscere e di agire in base a ragioni per l'azione che sono indipendenti dai desideri. Esplorerò questa e altre caratteristiche della razionalità umana nel corso di questo libro.

P75

2.

LA STRUTTURA DI BASE DELL'INTENZIONALITÀ, DELL'AZIONE E DEL SIGNIFICATO

Nel capitolo 1 ho affermato che molti degli errori presenti nella discussione della ragione pratica derivano da una concezione errata della razionalità, dalla concezione che ho chiamato il “Modello Classico”. Ma c'è una seconda ragione che provoca un certo numero di errori: gli autori in questione raramente partono da un'adeguata filosofia dell'intenzionalità e dell'azione. Tentare di scrivere sulla razionalità senza un'adeguata concezione generale della mente, del linguaggio e dell'azione è come provare a scrivere sui mezzi di trasporto senza conoscere nulla delle automobili, dei bus, dei treni e degli aerei. Per esempio, una domanda che viene posta comunemente è la seguente: che cosa sta all'azione nello stesso modo in cui la verità sta alla credenza? L'idea è che se noi potessimo fare maggiore chiarezza riguardo agli obiettivi dell'azione nello stesso modo in cui facciamo chiarezza riguardo alla relazione tra credenza e verità, allora, in un modo o nell'altro, l'argomento della ragione pratica diventerebbe più chiaro. Ma l'intera questione è confusa. Non vi è nulla che stia all'azione così come la credenza sta alla verità, per delle ragioni che spero diverranno completamente chiare quando avrò spiegato la struttura intenzionale delle azioni.

In questo capitolo presenterò, con un'esposizione semplice, una teoria generale della struttura intenzionale dell'azione umana, del significato e dei fatti istituzionali. È impossibile comprendere l'azione razionale se non si comprende, in primo luogo, che cos'è un'azione intenzionale, ed è impossibile comprendere le ragioni per l'azione se non si comprende il modo in cui gli esseri umani possono creare dei vincoli e altre entità dotate di significato e quindi creare delle ragioni. Ma è impossibile comprendere queste nozioni senza prima avere una qualche comprensione dell'intenzionalità in generale. Fintantoché al lettore non siano chiare alcune nozioni di base come il modo psicologico, il contenuto intenzionale, le condizioni di soddisfazione, la direzione di adattamento, la causazione intenzionale, l'auto-referenzialità causale, le funzioni di *status*, ecc., egli non capirà l'argomentazione che

segue. Ciò che dirò in questo capitolo è quasi interamente una ripetizione di quanto ho esposto in altri miei libri, specialmente in *Dell'intenzionalità*¹ e ne *La costruzione della realtà sociale*.² Per una esposizione più dettagliata dei punti affrontati in questo capitolo, così come per gli argomenti a supporto di queste conclusioni, il lettore dovrebbe consultare questi libri. I lettori che conoscono gli argomenti trattati in questi libri possono leggere questo capitolo con una certa rapidità.

Non so come presentare il materiale di questo capitolo in maniera efficace, se non esponendolo secondo lo stile del *Tractatus*³, ovvero come un insieme di proposizioni numerate.

1. *La definizione dell'Intenzionalità: l'Intenzionalità è direzionalità.*

“L'intenzionalità”, secondo l'uso della parola che ne fanno i filosofi, si riferisce a quell'aspetto degli stati mentali per il quale essi sono diretti a, o riguardano, o appartengono a stati di cose nel mondo, che vanno oltre essi stessi. L'“Intenzionalità” non ha una speciale connessione con l'“intendere”, nel senso ordinario della lingua italiana, come nel caso in cui affermo, per esempio, che “intendo andare al cinema stasera”. L'intendere (l'avere intenzione di) è solamente un tipo di intenzionalità tra le altre. Così, per esempio, le credenze, le paure, le speranze, i desideri e le intenzioni sono tutti stati intenzionali, allo stesso modo in cui lo sono emozioni come l'amore e l'odio, la paura e la gioia, l'orgoglio e la vergogna. Ogni stato che è diretto verso qualcosa al di là di se stesso è uno stato intenzionale. Così, per esempio, le esperienze visive sono intenzionali, ma le ansie non dirette verso qualcosa di preciso non lo sono.

2. *Gli stati intenzionali sono costituiti da un contenuto e da un modo psicologico, e spesso il contenuto è un'intera proposizione.*

Gli stati intenzionali, caratteristicamente, hanno una struttura analoga a quella degli atti linguistici. Così come posso ordinarvi di lasciare la stanza, chiedervi se lascerete la stanza e predire che voi lascerete la stanza, allo stesso modo posso sperare che voi lascerete la stanza, temere che voi lascerete la stanza, o desiderare che voi lascerete la stanza. In ogni

¹ John R. Searle, *Dell'intenzionalità. Un saggio di filosofia della conoscenza.*, tr. it. Bompiani, Milano 1985.

² John R. Searle, *La costruzione della realtà sociale*, tr. it. Edizioni di Comunità, Milano 1996.

³ *N.d.t.*, L'autore si riferisce al *Tractatus logico-philosophicus* (tr.it. Einaudi, Torino, 1961) di Ludwig Wittgenstein, opera strutturata secondo una sofisticata elencazione numerica delle proposizioni che la compongono.

caso c'è un contenuto proposizionale, che voi lascerete la stanza, ed esso si presenta in uno dei diversi e numerosi modi linguistici o psicologici. Nel caso del linguaggio esso può, per esempio, presentarsi nella forma di una domanda, di una predizione, di una promessa o di un ordine. Nel caso della mente, esso può, presentarsi in forma di credenze, paure e desideri. Per questa ragione rappresenterò la struttura generale dell'intenzionalità in questa forma:

$S(p)$

La “S” in questa formula indica il tipo di stato psicologico e la “p” indica il contenuto proposizionale dello stato. È essenziale porre questa distinzione perché lo stesso contenuto proposizionale può ricorrere in modi psicologici differenti. Per esempio, io posso sia credere che piovenga che sperare che piovenga; e, naturalmente, lo stesso modo psicologico, come una credenza, può comprendere un infinito numero di differenti contenuti proposizionali. Posso credere a qualsiasi genere di cose.

Non tutti gli stati intenzionali hanno come contenuto intenzionale un'intera proposizione. Le credenze e i desideri hanno delle intere proposizioni, ma non necessariamente le hanno l'amore e l'odio. Una persona, per esempio, può semplicemente amare Sally o odiare Harry. Per questa ragione, alcuni filosofi identificano gli stati intenzionali con un intero contenuto proposizionale come “attitudini proposizionali”. Credo che questa terminologia sia confusa, perché suggerisce che una credenza o un desiderio sono un'attitudine verso una proposizione, ma non mi sembra che le cose stiano veramente così. Se credo che Clinton sia il presidente, la mia attitudine è rivolta verso Clinton, l'uomo stesso, non verso la proposizione. La proposizione è il *contenuto*, non l'oggetto, della mia credenza. Pertanto, eviterò la terminologia “attitudini proposizionali”, mi riferirò soltanto agli stati intenzionali e opererò una distinzione all'interno degli stati intenzionali tra quelli che hanno delle intere proposizioni come contenuti e quelli che non le hanno. Così, la differenza tra il credere che Clinton sia il presidente e l'odiare Harry sarà rappresentata come segue:

Credenza (Clinton è il presidente)

Odio (Harry)

3. *Gli stati intenzionali proposizionali hanno delle condizioni di soddisfazione e una direzione di adattamento.*

Gli stati intenzionali con un contenuto proposizionale si possono adattare o meno alla realtà, e il modo in cui si suppone che essi si adattino alla realtà è determinato dal modo psicologico. Le credenze, per esempio, sono vere o false, a seconda che il contenuto della credenza si adatti o meno a una realtà che esiste in modo indipendente. Ma i desideri non sono né veri né falsi; essi vengono esauditi o frustrati se la realtà si adatta, o meno, al contenuto del desiderio. Le intenzioni, come i desideri, non sono né vere né false ma vengono realizzate o meno se il comportamento della persona che ha l'intenzione si adatta, oppure no, al contenuto dell'intenzione. Per fornire una spiegazione per questi fatti abbiamo bisogno delle nozioni di *condizioni di soddisfazione e direzione di adattamento*. Gli stati intenzionali come le credenze, i desideri e le intenzioni hanno condizioni di soddisfazione e direzioni di adattamento. Una credenza è soddisfatta se è vera, non soddisfatta se è falsa. Un desiderio sarà soddisfatto se esaudito, non soddisfatto se frustrato. Un'intenzione sarà soddisfatta se verrà realizzata, non soddisfatta se non verrà realizzata.

Inoltre, queste condizioni di soddisfazione sono rappresentate da differenti *direzioni di adattamento*, o da differenti responsabilità riguardo all'adattamento. Così, per esempio, una credenza può essere vera o falsa a seconda che il contenuto proposizionale della credenza si adatti realmente, o meno, al modo in cui le cose sono nel mondo, che esiste indipendentemente dalla credenza. Per esempio, se credo che stia piovendo la mia credenza sarà vera, e quindi soddisfatta, se e solo se sta piovendo. Poiché è responsabilità della credenza adattarsi a uno stato di cose del mondo, che esiste indipendentemente, possiamo dire che la credenza ha una *direzione di adattamento mente-a-mondo*. È compito della credenza, in quanto parte della mente, rappresentare o adattarsi a una realtà che esiste indipendentemente, ed avrà successo o fallirà a seconda che il contenuto mentale della credenza si adatti effettivamente alla realtà del mondo. I desideri, d'altro canto, hanno una direzione di adattamento opposta a quella delle credenze. I desideri non rappresentano il modo in cui le cose sono nel mondo, ma come vorremmo che fossero. Potremmo dire che è compito del mondo adattarsi al desiderio. I desideri e le intenzioni, diversamente dalle credenze, hanno la *direzione di adattamento mondo-a-mente*. Se la mia credenza è falsa io posso modificarla cambiando la credenza, ma se il mio desiderio non è soddisfatto, io non

posso, allo stesso modo, sistemare le cose cambiando il desiderio. Per modificare le cose, è il mondo che deve cambiare per adattarsi al contenuto del desiderio. Per questa ragione sostengo che i desideri e le intenzioni, diversamente dalle credenze, hanno una *direzione di adattamento mondo-a-mente*.

Questa distinzione, nel linguaggio ordinario è contrassegnata dal fatto che dei desideri e delle intenzioni non diciamo che sono veri o falsi. Diciamo piuttosto che il desiderio è esaudito o frustrato; e dell'intenzione che è realizzata oppure no, e questo dipende dal fatto che il mondo si adatti o meno al contenuto del desiderio o dell'intenzione. Il modo più semplice e diretto per verificare dunque se uno stato intenzionale ha una direzione di adattamento mente-a-mondo dipende dal fatto di poter affermare o meno se esso è vero o falso.

Alcuni stati intenzionali, così come molte emozioni, non hanno una direzione di adattamento in questo senso, poiché essi presuppongono che il contenuto proposizionale dell'emozione sia già soddisfatto. Così, se sono felicissimo che la Francia ha vinto la coppa del mondo, semplicemente do per scontato che la Francia abbia vinto la coppa del mondo. La mia gioia ha come contenuto proposizionale il fatto che la Francia ha vinto la coppa del mondo, ed io presuppongo che il contenuto proposizionale si adatti alla realtà. Non si tratta del fatto che lo stato intenzionale rappresenti sia come io credo che il mondo sia realmente, sia come vorrei che esso fosse; si presuppone piuttosto che il contenuto proposizionale si adatti alla realtà. In questi casi, sostengo che lo stato intenzionale ha una direzione di adattamento nulla o zero. Possiamo allora identificare tre direzioni di adattamento: mente-a-mondo, che è una caratteristica delle credenze e di altri stati cognitivi; mondo-a-mente, che è una caratteristica delle intenzioni e dei desideri, così come di altri stati volitivi e conativi; e la direzione di adattamento nulla, che è una caratteristica delle emozioni, come l'orgoglio e la vergogna, la gioia e la disperazione. Sebbene molte emozioni non abbiano una direzione di adattamento in quanto tale, esse contengono desideri e credenze, e queste hanno invece direzioni di adattamento. Così le emozioni come l'amore e l'odio possono giocare un ruolo nella ragione pratica perché essi contengono desideri, e questi desideri hanno una direzione di adattamento e possono quindi motivare delle azioni razionali. Questa caratteristica si rivelerà importante nell'ambito della nostra discussione sulla motivazione.

Le nozioni di condizioni di soddisfazione e di direzione di adattamento si applicano sia alle entità mentali che a quelle linguistiche. In effetti, è a partire dal parallelismo con gli atti linguistici che sono giunto a molte delle conclusioni che riguardano la natura della

mente. Gli enunciati, similmente alle credenze, rappresentano le loro condizioni di soddisfazione con la direzione di adattamento parola-a-mondo (analoga a quella di mente-a-mondo); ordini e promesse, analogamente ai desideri e alle intenzioni, rappresentano le loro condizioni di soddisfazione con la direzione di adattamento mondo-a-parola (analoga a quella di mondo-a-mente).

4. *Nel mondo vi sono molte entità che sebbene non siano nozioni mentali o linguistiche, in senso stretto, hanno condizioni di soddisfazione e direzione di adattamento*

La mappa di un territorio, per esempio, può essere accurata o non accurata; ha la direzione di adattamento mappa-a-mondo. Il progetto per una casa in costruzione potrà essere seguito oppure no; esso ha la direzione di adattamento mondo-a-progetto. Si suppone che l'appaltatore costruisca un edificio che si adatti al progetto. Nemmeno i bisogni, gli obblighi, le richieste e i doveri sono delle entità linguistiche, in senso stretto, ma anch'essi hanno dei contenuti proposizionali e delle direzioni di adattamento. Essi presentano le stesse direzioni di adattamento dei desideri, delle intenzioni, degli ordini e delle promesse. Se, per esempio, mi trovo costretto a pagare del denaro, allora il mio obbligo tale obbligo verrà assolto (soddisfatto) se e solo se pagherò quel denaro. Così l'obbligo è soddisfatto se e solo se il mondo cambia fino ad adattarsi al contenuto dell'obbligo. Analogamente agli obblighi, i bisogni, le richieste, gli impegni e i doveri, hanno una direzione di adattamento che richiede, perché essi vengano soddisfatti, che il mondo cambi per adattarsi al bisogno, alla richiesta, all'impegno o al dovere.

Mi piace usare delle metafore molto semplici e immaginare che fenomeni come le credenze, gli enunciati e le mappe siano degli enti sospesi sopra il mondo, che indicano in basso verso il mondo che essi rappresentano. Così penso alle direzioni di adattamento linguaggio-a-mondo e mente-a-mondo come dirette verso il basso. E talvolta rappresento queste direzioni di adattamento con una freccia rivolta all'ingiù. Analogamente i desideri, le intenzioni, gli ordini, le promesse, gli obblighi e gli impegni hanno le direzioni di adattamento mondo-a-mente e mondo-a-linguaggio. Penso a queste direzioni di adattamento come rivolte verso l'alto e le rappresento con una freccia che va all'insù. Per evitare locuzioni ingombranti talvolta dirò semplicemente “verso il basso” e “verso l'alto” o, talvolta, semplicemente disegnerò una freccia all'ingiù o all'insù.

Non posso dare troppa importanza a questa discussione piuttosto arida ai fini della comprensione dell'intenzionalità. *La chiave per comprendere la razionalità nell'azione è comprendere le relazioni tra il divario (gap) e la direzione di adattamento verso l'alto.*

5. *Gli stati intenzionali spesso funzionano causalmente tramite un tipo speciale di causazione, la causazione intenzionale, e in alcuni di essi la causazione è costruita all'interno delle loro condizioni di soddisfazione. Questi stati sono causalmente auto-referenziali.*

La nozione generale di causazione esprime il fatto che un qualcosa fa sì che qualcos'altro accada. Così, in un esempio classico, la palla da biliardo A colpisce la palla da biliardo B, causandone il movimento. Si è sostenuto che questo tipo di causazione è soltanto uno dei diversi tipi di causazione: quella che, da Aristotele in poi, è definita la "causa efficiente". Continuando a usare la terminologia di Aristotele, si suppone vi siano almeno altri tre tipi di causazione: la causa formale, la causa finale e la causa materiale. Io credo che tutto questo sia molto confuso. Esiste un solo tipo di causazione e questa è la causa efficiente. Tuttavia, all'interno della causa efficiente, c'è un'importante sottocategoria che ha a che fare con la causazione mentale. Si tratta di casi in cui qualcosa causa uno stato mentale, o in cui uno stato mentale causa qualcos'altro. E all'interno della sottocategoria della causazione mentale esiste ancora un'altra categoria, quella della *causazione intenzionale*. Nel caso della causazione intenzionale, accade sia che uno stato intenzionale causi le proprie condizioni di soddisfazione, sia che le condizioni di soddisfazione causino lo stato intenzionale stesso. Per esporre questo punto in una terminologia leggermente differente, possiamo dire che nel caso della causazione intenzionale, lo stato intenzionale causa l'effettivo stato di cose che esso rappresenta, o lo stato di cose che esso rappresenta causa lo stato intenzionale. Così, se voglio bere dell'acqua, il mio desiderio di bere dell'acqua causerà il fatto che io bevo dell'acqua, ed avrò così un caso di causazione intenzionale. Il desiderio ha come contenuto il fatto che *io bevo dell'acqua* e quel desiderio fa sì che *io beva dell'acqua* (sebbene dobbiamo ricordare che in tali casi di azione volontaria esiste generalmente un divario). Se vedo che il gatto è sulla stuoia, allora il fatto che il gatto è sulla stuoia causa la reale esperienza visiva, e parte delle condizioni di soddisfazione di quest'ultima sono date dal fatto che il gatto è sulla stuoia. La causazione intenzionale è una qualsiasi relazione causale tra uno stato intenzionale e le sue condizioni

di soddisfazione, dove lo stato intenzionale causa le sue condizioni di soddisfazione o le sue condizioni di soddisfazione causano lo stato intenzionale stesso.

Così come abbiamo scoperto che la nozione di *direzione di adattamento* è essenziale per comprendere i modi in cui l'intenzionalità e il mondo reale entrano in relazione tra loro, allo stesso modo mi sembra che abbiamo bisogno della nozione di *direzione di causazione*. Se ho sete e bevo dell'acqua per soddisfare la mia sete, allora la mia sete, essendo tra le altre cose, un desiderio di bere acqua, avrà una direzione di adattamento mondo-a-mente (verso l'alto). Il desiderio di bere, se soddisfatto, sarà soddisfatto da un cambiamento del mondo in modo che il mondo si adatti al contenuto del desiderio: il mio desiderio di bere dell'acqua. Ma se il mio desiderio causa il fatto che io beva dell'acqua allora la relazione causale tra il mio desiderio e il fatto che io beva è del tipo mente-a-mondo. Il desiderio che si trova nella mia mente causa il fatto che io (tenendo presente il divario, naturalmente) beva dell'acqua, all'interno del mondo. In questo caso, la direzione di adattamento mondo-a-mente è parallela rispetto alla direzione di causazione mente-a-mondo. Nel caso della percezione visiva, per esempio, la direzione di adattamento e la direzione di causazione sono differenti. Se l'esperienza visiva è, come si suol dire, *veritiera*, allora l'esperienza visiva si adatterà al mondo, e avremo una direzione di adattamento mente-a-mondo ben riuscita. Ma se l'esperienza visiva è veramente soddisfatta, siamo di fronte al caso in cui lo stato di cose del mondo che percepisco causa la reale esperienza visiva attraverso la quale io percepisco quello stato di cose. Così, in questo caso, la direzione di adattamento mente-a-mondo è parallela alla direzione di causazione mondo-a-mente.

Questo esempio illustra una speciale sottoclasse di casi di causazione intenzionale, in cui è parte delle condizioni di soddisfazione dello stato intenzionale in questione la necessità che esso stesso funzioni casualmente nel produrre le sue condizioni di soddisfazione, perché venga soddisfatto. Così, nel caso delle intenzioni, diversamente che nei desideri, l'intenzione non è realmente realizzata fino a che l'intenzione stessa non causi l'azione che è veramente rappresentata nel contenuto dell'intenzione. Se l'azione ha una causa differente allora l'intenzione non è realizzata. In tali casi si potrebbe allora dire che le condizioni di soddisfazione dello stato intenzionale sono *casualmente autoreferenziali*⁴. I casi di stati intenzionali che sono casualmente autoreferenziali sono: le esperienze

⁴ L'individuazione del fenomeno dell'autoreferenzialità causale risale a molto tempo fa. Fu messo in rilievo per esempio da Kant all'interno della sua discussione sulla causalità della volontà. Questa terminologia, a quanto mi risulta, è piuttosto recente ed è stata usata per la prima volta da Gilbert Barman in "Practical Reasoning", *Review of Metaphysics*, n. 29, 1976, pp.431-463.

percettive, i ricordi e le intenzioni. Consideriamo ognuno di questi, secondo l'ordine. Nel caso dell'esperienza percettiva, l'esperienza sarà soddisfatta soltanto se il reale stato di cose che è percepito intenzionalmente causa quell'effettiva esperienza percettiva. Così, per esempio, se vedo che il gatto è sulla stuoia, il contenuto intenzionale dell'esperienza visiva sarà:

Esp. Vis. (che il gatto è sulla stuoia, e il fatto che il gatto è sulla stuoia causa questa Esp. Vis.).

Questa formula deve essere letta nel modo seguente: in questo momento io ho un'esperienza visiva le cui condizioni di soddisfazione sono che il gatto è sulla stuoia, e il fatto che il gatto è sulla stuoia causa questa esperienza visiva. Notate come sia necessario distinguere tra ciò che viene realmente visto e le condizioni di soddisfazione totali dell'esperienza visiva. Ciò che viene realmente visto è il fatto che il gatto è sulla stuoia, ma le condizioni totali di soddisfazione dell'esperienza visiva includono una componente causalmente autoreferenziale. È importante sottolineare che io non vedo effettivamente la causazione – io vedo un gatto e una stuoia, e vedo il primo sulla seconda. Ma affinché io possa essere capace di fare questo, deve esserci una componente causale relativa alle condizioni totali di soddisfazione dell'esperienza visiva, ed è proprio questa caratteristica logica che ho cercato di fissare nella formula sopra esposta.

In modo simile, anche i ricordi sono causalmente autoreferenziali. Se ricordo che ieri ero andato ad un picnic, allora le condizioni di soddisfazione sono sia il fatto che ieri sono andato ad un picnic, sia il fatto che l'essere andato al picnic ieri causa proprio questo ricordo. Notate che nel caso della percezione e del ricordo noi abbiamo una direzione di adattamento mente-a-mondo e una direzione di causazione mondo-a-mente. Sia nel caso della percezione che in quello del ricordo, se osservo il modo in cui il mondo è realmente, o se ricordo com'era, e realizzo dunque una direzione di adattamento mente-a-mondo, è solo perché il fatto che il mondo sia in quel modo, o che sia stato in quel modo, causa la mia esperienza percettiva e il mio ricordo, consentendomi di realizzare la direzione di causazione mondo-a-mente. Si può ottenere una direzione di adattamento mente-a-mondo solo grazie alla realizzazione di una direzione di causazione mondo-a-mente.

Troviamo l'autoreferenzialità causale anche all'interno della struttura dell'intenzione e dell'azione. Ecco, con un esempio molto semplice, come essa funziona. Io dispongo di un

insieme di credenze e di desideri e ragionando su queste credenze e su questi desideri arrivo a formare un'intenzione. Chiamerò queste intenzioni che si formano precedentemente all'azione, *intenzioni precedenti*. Supponiamo, per esempio, che durante una riunione io voglia votare per una mozione che è stata proposta, e credo di poter votare per questa mozione sollevando il mio braccio destro. Formo così l'intenzione precedente di sollevare il mio braccio. Il contenuto intenzionale dell'intenzione precedente di sollevare il braccio può essere rappresentato in questo modo:

i.p. (che io sollevo il braccio e che questa *i.p.* causa il fatto che io sollevo il braccio).

Questa formula deve essere letta nel modo seguente: io ho un'intenzione precedente le cui condizioni di soddisfazione sono che io sollevi il mio braccio, e proprio questa intenzione precedente causa il fatto che io sollevi il mio braccio.

L'intenzione precedente deve essere distinta da quella che chiamo *intenzione-in-azione*. L'intenzione-in-azione è l'intenzione che ho mentre sto effettivamente compiendo un'azione. Quindi, in questo caso, quando arriva il momento di votare e il presidente dice: "Chi è a favore alzi il braccio", io agirò in base alla mia intenzione precedente e così avrò un'intenzione in azione le cui condizioni di soddisfazione sono che proprio quella intenzione-in-azione dovrebbe causare il movimento corporeo del sollevare il mio braccio. Possiamo rappresentare tutto questo nel modo seguente:

i.a. (il mio braccio si solleva e questa *i.a.* causa il fatto che il mio braccio si solleva).

Questa formula deve essere letta nel modo seguente: io ho un'intenzione-in-azione le cui condizioni di soddisfazione sono il fatto che il mio braccio si sollevi e che proprio questa intenzione-in-azione causa il fatto che il mio braccio si sollevi.

In italiano il termine più vicino all'intenzione-in-azione è il "tentare". Se avevate un'intenzione-in-azione ma non siete riusciti a realizzare le sue condizioni di soddisfazione, avrete al massimo tentato. In un caso tipico di azione premeditata in cui io agisco sulla base di un'intenzione precedente, come quando sollevo il mio braccio, la struttura di tutto questo è che per prima cosa mi sono formato un'intenzione precedente (le cui condizioni di soddisfazione sono il fatto che essa dovrebbe causare l'intera azione) e, in seguito, ho compiuto l'intera azione, dove l'intera azione consiste di due componenti: l'intenzione-in-

azione e il movimento corporeo (e la condizione di soddisfazione dell'intenzione-in-azione è che essa dovrebbe causare il movimento corporeo).

Certamente, non tutte le azioni sono premeditate. Molte delle cose che faccio, le faccio in maniera spontanea. In questo caso io ho un'intenzione-in-azione, ma non un'intenzione precedente. Per esempio, talvolta, quando penso a un problema filosofico, mi alzo in piedi e cammino per la stanza. Il mio camminare per la stanza è compiuto intenzionalmente, anche se non ho alcuna intenzione precedente. I miei movimenti corporei sono in questo caso causati da un'intenzione-in-azione continua, ma non vi è alcuna intenzione precedente.

6. *Le strutture intenzionali della cognizione e della volizione sono immagini speculari l'una dell'altra, con direzioni di adattamento e direzioni di causazione che vanno in direzioni opposte.*

Se iniziamo con il prendere in considerazione l'azione e la percezione possiamo osservare le seguenti simmetrie e le asimmetrie. Le percezioni consistono di due componenti. Nel caso della visione, per esempio, una percezione consiste di un'esperienza visiva cosciente assieme ad uno stato di cose percepito. Così se vedo che il gatto è sulla stuoia, allora ho anche l'esperienza visiva e c'è un corrispondente stato di cose nel mondo, cioè che il gatto è sulla stuoia. Inoltre, se l'esperienza visiva è soddisfatta, la sua componente causalmente autoreferenziale deve essere soddisfatta: lo stato delle cose nel mondo che sto percependo deve causare l'esperienza effettiva del percepire. L'azione umana è esattamente parallela ma con opposte direzioni di adattamento e di causazione. Così un'azione intenzionale compiuta con successo consiste di due componenti: un'intenzione-in-azione e, solitamente, un movimento corporeo. Se io sollevo il mio braccio nell'esecuzione di un'azione umana si ha un'intenzione-in-azione; ed essa ha come condizioni di soddisfazione il fatto che il mio braccio si solleva e che proprio quella intenzione-in-azione causa il fatto che il mio braccio si solleva. Così le due componenti dell'azione intenzionale eseguita con successo sono l'intenzione-in-azione e il movimento corporeo.

Le simmetrie e le asimmetrie delle relazioni tra la percezione e l'azione sono tipiche della cognizione e della volizione in generale. Abbiamo osservato prima che gli stati cognitivi della percezione e della memoria hanno una direzione di adattamento mente-a-mondo e una direzione di causazione mondo-a-mente. Ma l'intenzione precedente e l'intenzione-in-azione hanno direzioni di adattamento e di causazione opposte. Esse hanno

direzione di adattamento mondo-a-mente e direzione di causazione mente-a-mondo. Questo è solo un altro modo per dire che l'intenzione viene realizzata solo se il mondo diventa come l'intenzione se lo rappresenta e l'intenzione fa sì che ciò accada. Così, per essere soddisfatta, l'intenzione deve produrre una direzione di adattamento mondo-a-mente e una direzione di causazione mente-a-mondo. L'intenzione sarà soddisfatta solo se l'intenzione stessa funziona causalmente nel realizzare una direzione di adattamento mondo-a-mente. In tal caso noi realizziamo una direzione di adattamento verso l'alto solo in virtù di una direzione di causazione verso il basso. Un tipico modello, di azione premeditata è quello secondo il quale un'intenzione precedente si forma sulla base di credenze e desideri. L'intenzione precedente è una rappresentazione di un'azione intera e l'azione intera consiste di due componenti – l'intenzione-in-azione e il movimento corporeo. Se l'intenzione

precedente viene portata a termine essa causerà l'intenzione-in-azione la quale, a sua volta, causerà il movimento corporeo. L'intera struttura formale delle relazioni tra la cognizione e la volizione è descritta nella tabella 1.

Tabella 1

	Cognizio ne			Volizione		
	Credenz a	Memori a	Percezion e	Desiderio	Intenzion e preceden te	Intenzion e-in- azione
Direzione adattamento	di ↓	↓	↓	↑	↑	↑
Direzione di causazione determinata dalle condizioni di soddisfazione	nessuna	↑	↑	nessuna	↓	↓
Causalmente autoreferenziali	No	Sì	Sì	No	Sì	Sì

Le intenzioni in azione possono essere coscienti oppure no. Quando sono esperienze coscienti le chiamerò “esperienze dell’agire” e credo che ciò che io definisco esperienza dell’agire sia quella che William James definiva come la sensazione dello “sforzo”.⁵

7. *La deliberazione porta all’azione intenzionale tramite le intenzioni precedenti.*

In un caso semplice dove le sole ragioni sono le credenze e i desideri possiamo dire che: la riflessione operata sulle credenze e sui desideri, con le loro differenti direzioni di adattamento, conduce a una decisione, cioè, alla formazione di un’intenzione precedente che ha una direzione di adattamento verso l’alto e una direzione di causazione verso il basso. L’intenzione precedente ha la condizione di soddisfazione che causa l’azione. L’azione consiste di due componenti, l’intenzione-in-azione e il movimento corporeo, e l’intenzione-in-azione ha come sua condizione di soddisfazione il fatto di causare il movimento corporeo. Così la sequenza nel caso dell’azione premeditata è la seguente:

La deliberazione causa l’intenzione precedente, la quale causa l’intenzione-in-azione, la quale, a sua volta, causa il movimento corporeo. L’azione totale consiste nell’intenzione-in-azione e nel movimento corporeo. Il modello, allora, può essere rappresentato in questo modo, con le frecce che indicano la relazione causale:

Deliberazione sulla base di credenze e desideri → intenzione precedente → intenzione-in-azione → movimento corporeo (azione = intenzione-in-azione + movimento corporeo)

Nel caso della volizione, la direzione di adattamento degli stati causalmente autoreferenziali è sempre mondo-a-mente e la direzione di causazione mente-a-mondo. Nel caso della cognizione, la direzione di adattamento degli stati causalmente autoreferenziali è sempre mente-a-mondo e la direzione di causazione sempre mondo-a-mente. L’intenzione verrà soddisfatta, e in questo modo si realizzerà la direzione di adattamento mondo-a-mente, soltanto se l’intenzione stessa funziona causalmente in modo da provocare quell’adattamento. Le percezioni e i ricordi saranno soddisfatti, e si realizzerà una direzione di adattamento mente-a-mondo, solo se il mondo stesso causa proprio quelle percezioni e

⁵ William James, *The Principles of Psychology*, Volume II, Chapter 26, New York, Henry Holt, 1918, tr. it. William James, *Principi di psicologia*, Principato, Milano 1999, vol II, cap. 26.

quei ricordi. Disponiamo dunque di una direzione di adattamento mente-a-mondo soltanto in virtù della direzione di causazione mondo-a-mente.

8. *La struttura della volizione contiene tre divari.*

Una volta tenuto conto delle differenze tra la direzione di adattamento e la direzione di causazione, l'asimmetria principale tra la struttura formale della cognizione da un lato, e la volizione da un altro, è costituita dal fatto che la volizione contiene dei divari. "Divario" è il nome generale che ho introdotto per descrivere il fenomeno per il quale noi, normalmente, *non* abbiamo un'esperienza dei diversi stadi delle nostre deliberazioni e delle nostre azioni volontarie come se esse avessero condizioni causalmente sufficienti per passare allo stadio successivo. Per i propositi di questo libro, possiamo segmentare l'esperienza continua del divario come segue. Nella struttura della deliberazione e dell'azione c'è, innanzitutto, un divario tra le deliberazioni e le intenzioni precedenti che sono il risultato delle deliberazioni. Così, quando io decido di votare o di non votare la mozione, c'è un divario tra le ragioni che ho a favore o contro quella mozione e la decisione effettiva, ovvero la formazione effettiva di un'intenzione precedente di votare per quella mozione. Inoltre c'è un divario tra l'intenzione precedente e l'intenzione-in-azione, cioè il divario tra il decidere di fare qualcosa e il reale tentativo di farlo. Non esiste un divario del genere tra l'intenzione-in-azione e il movimento corporeo. Se sto effettivamente provando a fare qualcosa, e se ci riesco, il mio tentativo deve risultare causalmente sufficiente per la realizzazione dell'azione. Il terzo divario si trova all'interno della struttura delle intenzioni-in-azione che si estendono temporalmente. Quando ho un'intenzione-in-azione di impegnarmi in qualche tipo di attività complessa come scrivere un libro o attraversare a nuoto la Manica, il cominciamento della mia originale intenzione-in-azione non è di per sé sufficiente a garantire la continuazione di quella intenzione-in-azione fino al completamento dell'attività. Così in ogni fase del compimento di un'intenzione-in-azione c'è un terzo divario. Inoltre se ci troviamo di fronte ad una qualche azione prolungata come attraversare a nuoto la Manica o scrivere un libro, la mia intenzione precedente continua ad essere causalmente effettiva per l'intera operazione. Cioè, devo continuare a compiere uno sforzo per portare a compimento il tipo di azione che avevo originariamente pianificato nell'ambito della formazione dell'intenzione precedente.⁶

⁶ Non mi ero accorto di questo punto quando ho scritto *Dell'intenzionalità*. In quel libro rappresento l'intenzione precedente come se essa finisse laddove l'intenzione-in-azione comincia. Ma questo è un errore.

9. *Le azioni complesse hanno una struttura interna secondo la quale l'agente intende fare una cosa intendendo farne un'altra, o intende fare una cosa facendone un'altra. Queste due relazioni sono, rispettivamente, causali e costitutive.*

Finora ho parlato come se, in effetti, si compissero soltanto delle azioni semplici. Ma, se si fa eccezione per azioni semplici come il sollevare un braccio, le azioni umane sono più complesse ed hanno una struttura interna complessa. Normalmente una persona fa una cosa facendo qualcos'altro o attraverso il fare qualcos'altro. Si accende la luce premendo l'interruttore, si spara con la pistola facendo pressione sul grilletto, per esempio. Anche nel semplice esempio che ho fornito, si vota sollevando il braccio, cioè attraverso il sollevamento del braccio. Non ci sono due azioni, il sollevare il braccio e il votare, ma solamente una singola azione: il votare sollevando braccio. La struttura interna dell'azione è molto importante per la comprensione della ragione pratica, poiché spesso la decisione riguarda la scelta tra il fare qualcosa tramite un'altra o il fare qualcosa facendone un'altra per raggiungere il proprio scopo. Nell'esempio della scimmia di cui abbiamo discusso nel capitolo 1, la scimmia prendeva le banane tramite l'azione di colpirla con un bastone. Le due più importanti forme strutturali della struttura interna delle azioni sono la relazione causale del tramite-qualcosa e la relazione costitutiva del facendo-un'altra-cosa. Se io sparo con la pistola tramite la pressione sul grilletto, il tipo di relazione è causale. Premere il grilletto causa il fatto che il colpo venga esploso. Se voto sollevando il mio braccio, la relazione è costitutiva. In quel contesto, il sollevare il braccio costituisce l'azione stessa di votare. Nel caso della relazione di tramite-qualcosa, la relazione tra le componenti dell'azione è una relazione di causazione. Il premere l'interruttore causa l'accensione della luce e quando ho acceso la luce tramite l'azione di premere l'interruttore, avevo un'intenzione-in-azione complessa, cioè avevo questa intenzione-in-azione che avrebbe dovuto causare il premere l'interruttore che, a sua volta, avrebbe causato l'accensione della luce. Ma quando ho sollevato il braccio per votare, il sollevarsi del mio braccio non ha causato il mio votare; piuttosto, il sollevarsi del mio braccio *ha costituito* il mio votare. In quel contesto il movimento corporeo costituisce, o equivale, all'azione stessa di votare. Nel caso delle azioni complesse, che si estendono per un lungo periodo di tempo, queste relazioni diventano piuttosto complesse. Consideriamo lo scrivere questo libro. Io ci lavoro

L'intenzione precedente può continuare ad essere effettiva per tutta l'esecuzione dell'azione. Questo errore mi è stato fatto notare da Brian O'Shaughnessy.

sedendomi al computer e digitando sulla tastiera i miei pensieri. Queste azioni non causano la scrittura del libro, ma sono costitutive delle varie fasi. Quando premo i tasti del computer, d'altro canto, le mie azioni causano il fatto che il testo del libro appaia sullo schermo.

Un'altra idealizzazione che ho impiegato è stata quella di parlare come se tutte le azioni fossero casi di intenzioni-in-azione che causano dei movimenti corporei. Ma, naturalmente, vi sono anche le azioni mentali, per esempio, che accadono nella mente di ciascuno. E vi sono azioni negative, per esempio l'astenersi dal fumare. Ci sono anche quelle azioni, come ho menzionato sopra, che sono estese nel tempo, come lo scrivere un libro o l'allenarsi per una gara di sci. Credo che la spiegazione che ho dato con la distinzione tra intenzioni precedenti e intenzioni-in-azione, e la distinzione, nella struttura interna, tra le relazioni di tramite-qualcosa e le relazioni costitutive del facendo-un'altra-cosa, fornisca una spiegazione per tutti questi casi.

10. *Il significato è una questione che riguarda l'imposizione intenzionale di condizioni di soddisfazione su altre condizioni di soddisfazione.*

Se, per esempio, un parlante dice "sta piovendo", e con questo enunciato vuole significare che sta piovendo, le condizioni di soddisfazione della sua intenzione-in-azione sono, prima di tutto, il fatto che l'intenzione-in-azione dovrebbe causare l'enunciazione della frase, "sta piovendo", e, in secondo luogo, l'enunciato stesso dovrebbe avere come condizione di soddisfazione, con direzione di adattamento verso il basso, il fatto che sta piovendo. Nel caso in cui il parlante esprime un significato, il parlante stesso crea una forma di intenzionalità imponendo intenzionalmente delle condizioni di soddisfazione su qualcosa che egli ha prodotto intenzionalmente, per esempio i suoni provenienti dalla sua bocca o i segni scritti sulla carta. Egli produce un enunciato intenzionalmente e produce un enunciato con l'intenzione ulteriore che quello stesso enunciato dovrebbe avere delle condizioni di soddisfazione.

Questo procedimento nel linguaggio naturale umano è reso possibile dal fatto che le parole all'interno delle frasi del linguaggio hanno una forma di intenzionalità che è essa stessa derivata dall'intenzionalità intrinseca, o indipendente dall'osservatore, degli agenti umani. E questo ci porta al mio prossimo punto.

11. *Abbiamo bisogno di distinguere tra intenzionalità indipendente dall'osservatore e intenzionalità dipendente dall'osservatore.*

Fino ad ora ho parlato dell'intenzionalità della mente umana. Ma è possibile ascrivere un'intenzionalità a cose diverse dalla mente umana, cose che sono letteralmente vere e in cui l'intenzionalità dipende dall'intenzionalità della mente intrinseca o indipendente dall'osservatore. Nel caso del linguaggio, risulta evidente che le parole e le frasi hanno un significato, e il significato è una forma di intenzionalità. Questa è la differenza tra l'affermare "ho fame", che letteralmente attribuisce intenzionalità a me stesso, e dire che "la frase in francese 'J'ai faim' significa che io ho fame". Attribuendo un significato alla frase, le ho attribuito una forma di intenzionalità. Ma l'intenzionalità della frase in francese non è, per così dire, intrinseca; essa è derivata dall'intenzionalità dei parlanti francesi. Così dirò che c'è una distinzione tra l'intenzionalità *indipendente dall'osservatore* del mio stato mentale di fame, e l'intenzionalità *dipendente dall'osservatore* o *relativa all'osservatore* delle parole e delle frasi in francese, in inglese e in altre lingue. C'è una terza forma di ascrizione di intenzionalità che non è né indipendente dall'osservatore né relativa all'osservatore, e che non è per niente letterale. Sto pensando a cose come l'attribuire un ricordo a un computer o un desiderio ad una pianta. Questo è un modo di esprimersi poco efficace. Se dico: "le mie piante sono assetate d'acqua", nessuno rimarrà confuso a pensare che io sto letteralmente attribuendo un'intenzionalità alle piante. Chiamerò queste ascrizioni di intenzionalità metaforiche o del "come-se". Tuttavia, non sto attribuendo un terzo tipo di intenzionalità; piuttosto, le piante, i computer e molte altre cose si comportano *come se* possedessero intenzionalità e così possiamo attribuire loro queste attribuzioni metaforiche, ascrizioni del come-se, sebbene essi non abbiano, letteralmente parlando, alcuna intenzionalità.

12. *La distinzione tra oggettività e soggettività è in realtà la fusione di due distinzioni: una ontologica e una epistemica*

Possiamo usare la distinzione tra forme di intenzionalità relative all'osservatore e indipendenti dall'osservatore per compiere un'ulteriore distinzione che risulta importante per l'argomento successivo di questo libro. La nozione di *oggettività* e il contrasto tra *oggettività* e *soggettività* compare ampiamente nella nostra cultura intellettuale. Noi cerchiamo verità scientifiche che siano "oggettive". Tuttavia, c'è una grande confusione per quanto riguarda queste nozioni, una confusione che dovremmo eliminare. Dobbiamo distinguere tra *oggettività e soggettività ontologica* da un lato, e *oggettività e soggettività epistemica* dall'altro. Alcuni esempi renderanno chiara questa distinzione. Quando affermo

di provare un dolore io attribuisco a me stesso un'esperienza soggettiva. Quest'esperienza soggettiva ha un'ontologia soggettiva perché essa esiste solamente quando viene esperita da un soggetto cosciente. Sotto questo aspetto, i dolori, il solletico e i pruriti differiscono dalle montagne, dalle molecole e dai ghiacciai; questo perché le montagne ecc. hanno un'esistenza oggettiva, o un'ontologia oggettiva. La distinzione tra soggettività e oggettività ontologiche è molto diversa dalla distinzione tra soggettività e oggettività epistemiche. Se io dico: "Rembrandt trascorse l'intera vita nei Paesi Bassi", questa frase è epistemicamente oggettiva dal momento che noi possiamo accertare la sua verità o la sua falsità senza riferirci alle attitudini e ai sentimenti degli osservatori. Ma se affermo che "Rembrandt fu il più grande pittore mai vissuto ad Amsterdam", ebbene, questa è, come si dice, una questione di opinione. L'enunciato è epistemicamente soggettivo perché la sua verità non può essere stabilita indipendentemente dalle attitudini soggettive degli ammiratori e dei detrattori dei lavori di Rembrandt e di altri pittori di Amsterdam. Possiamo dire, alla luce di questa distinzione, che tutti i fenomeni relativi all'osservatore contengono un elemento di soggettività ontologica. Il fatto che qualcosa significhi qualcosa, come nel caso di una frase in francese, dipende, in questo caso, dalle attitudini ontologicamente soggettive dei parlanti francesi. Ma, e questo è il punto cruciale, la soggettività ontologica non implica necessariamente una soggettività epistemica. Possiamo avere una conoscenza epistemicamente oggettiva circa i significati delle frasi in francese, e in altre lingue, anche se questi significati sono ontologicamente soggettivi. Questa distinzione si rivelerà cruciale per noi più tardi, quando scopriremo che molte delle caratteristiche del mondo che causano le azioni razionali sono, analogamente, ontologicamente soggettive ma epistemicamente oggettive.

13. *L'intenzionalità collettiva consente la creazione di fatti istituzionali. I fatti istituzionali sono creati in accordo con le regole costitutive della forma "X equivale a Y in C".*

L'intenzionalità non può essere soltanto individuale, del tipo "io intendo andare al cinema", ma essa è anche collettiva, del tipo "noi intendiamo andare al cinema". L'intenzionalità collettiva consente ai gruppi di persone di creare dei fatti istituzionali comuni, come quelli che hanno a che vedere con il denaro, la proprietà, i matrimoni, i governi e, soprattutto, il linguaggio. In tali casi l'esistenza dell'istituzione consente agli individui o ai gruppi di individui di imporre agli oggetti funzioni che gli oggetti non

potrebbero realizzare solamente in virtù della loro struttura fisica, ma solo grazie ad un riconoscimento collettivo dell'oggetto come dotato di un certo *status* e, con questo *status*, di una speciale funzione. Io chiamo queste funzioni, *funzioni di status* ed esse hanno la forma caratteristica di "X equivale a Y in C". Così una certa sequenza di parole equivale a una frase in inglese o in italiano, un certo pezzo di carta equivale, negli Stati Uniti, a un biglietto da dieci dollari, una certa posizione equivale allo scacco matto negli scacchi, una persona che soddisfa certe particolari condizioni equivale all'essere il presidente degli Stati Uniti. Queste funzioni di *status* differiscono dalle funzioni fisiche perché un oggetto come un cacciavite compie la sua funzione fisica in virtù della sua struttura fisica, mentre le frasi in inglese, lo scacco matto, il denaro e i presidenti possono compiere le loro funzioni solo se sono riconosciuti collettivamente come dotati un certo *status*, e aventi la funzione di *status* relativa.

La combinazione della realtà istituzionale, essa stessa creata dall'imposizione di funzioni di status in accordo con la regola costitutiva "X equivale a Y in C", con una speciale forma di funzioni di *status*, in particolare l'imposizione di significato, consente agli esseri umani di creare certe forme di ragioni per l'azione indipendenti dai desideri. Analizzeremo questo fenomeno in dettaglio nel capitolo 6. Ora voglio solamente sottolineare quanto segue. Abbiamo osservato che il significato riguarda l'imposizione di condizioni di soddisfazione su altre condizioni di soddisfazione (punto 10); e questo si connette al fatto che i fatti istituzionali sono creati all'interno di sistemi istituzionali tramite i quali un agente impone una funzione su un'entità, dove l'entità stessa non può compiere quella funzione senza un qualche tipo di accettazione collettiva o di riconoscimento collettivo di quella funzione. Questi due fattori, presi insieme, ci permettono di notare come, nel compiere un atto linguistico quale l'esprimere un'asserzione o il fare una promessa, il parlante crei un nuovo insieme di condizioni di soddisfazione, e questo nuovo insieme di condizioni di soddisfazione è il risultato della creazione di un fatto istituzionale, per esempio il fatto che il parlante abbia espresso un'asserzione, o fatto una promessa a chi lo ascolta.

14. *L'intenzionalità funziona soltanto per determinare le condizioni di soddisfazione su uno Sfondo di abilità pre-intenzionali o non-intenzionali.*

In aggiunta alla struttura intenzionale della cognizione e della volizione abbiamo bisogno di spiegare che l'intero sistema dell'intenzionalità, per il quale gli stati intenzionali determinano delle condizioni di soddisfazione, funziona soltanto su uno Sfondo di abilità, capacità, tendenze e disposizioni che possiedono sia gli esseri umani che gli animali, ma che non consistono di per sé in stati intenzionali. Affinché io possa formare l'intenzione di camminare per la stanza o di lavarmi i denti o di scrivere un libro, devo essere capace di camminare per la stanza, di lavarmi i denti e di scrivere un libro, o per lo meno devo presupporre di essere capace di fare queste cose. Ma le mie abilità non sono di per sé ulteriori stati intenzionali, sebbene le abilità siano capaci di generare gli stati intenzionali. Pensate alle mie abilità, alle mie capacità, alle mie tendenze e alle mie disposizioni, da un punto di vista ontologico, come ad un insieme di strutture cerebrali. Queste strutture cerebrali mi permettono di attivare il sistema dell'intenzionalità e di farlo funzionare, ma le capacità realizzate nelle strutture cerebrali non consistono esse stesse di stati intenzionali.

Lo Sfondo è importante per comprendere la struttura della razionalità sotto diversi aspetti che vanno anche al di là degli obiettivi di questo libro. I casi apparenti di relatività culturale della razionalità sono solitamente dovuti a differenti Sfondi culturali. La razionalità in quanto tale è universale. E a questo punto della trattazione vorrei solamente richiamare l'attenzione sul fatto che il sistema dell'intenzionalità non è, per così dire, pienamente intenzionale sotto tutti gli aspetti. In aggiunta al sistema dell'intenzionalità dobbiamo supporre che gli agenti abbiano un insieme di abilità che non sono esse stesse ulteriori stati intenzionali. E questi insiemi di abilità sono ciò che definisco, con la lettera maiuscola, "lo Sfondo".

15. *L'intenzionalità-con-la-z deve essere distinta dall'intensionalità-con-la-s*

L'intenzionalità-con-la-z è quella proprietà della mente, e, del linguaggio, secondo la quale gli stati mentali e gli atti linguistici riguardano o appartengono a oggetti e a stati di cose del mondo. L'intensionalità-con-la-s è quella proprietà degli enunciati, e di altri tipi di rappresentazioni, per la quale essi falliscono alcune prove per l'estensionalità. Le due prove migliori sono fornite dalla prova della sostituibilità, che riguarda le espressioni in cui è possibile sostituire il co-riferimento senza perdere o cambiare il valore di verità (questa viene talvolta chiamata Legge di Leibniz), e dalla prova della generalizzazione esistenziale. Per esempio, la frase "Edipo vuole sposare Giocasta" fallisce la prova della sostituibilità perché l'enunciato "Giocasta è identica a sua madre", non permette l'inferenza "Edipo vuole

sposare sua madre”. L’enunciato è intensionale rispetto alla sostituibilità. Gli enunciati che falliscono la prova della sostituibilità sono talvolta definiti *referenzialmente opachi*. La frase “Edipo cerca la città perduta di Atlantide” non consente l’inferenza esistenziale “Esiste la città perduta di Atlantide”, poiché Edipo potrebbe comunque cercarla anche se la cosa che egli sta cercando non esistesse. Così le frasi falliscono la prova della generalizzazione esistenziale. L’intensionalità è importante per l’argomento della ragione pratica perché, tra le altre ragioni, gli enunciati che riguardano le ragioni per l’azione sono tipicamente intensionali-con-la-s.

16. Conclusioni

Mi scuso con il lettore per l’aridità, così come per la velocità, della discussione. Questo apparato mi servirà nei capitoli seguenti e, sinceramente, non me la sentivo di dire ai miei lettori di andarsi prima a leggere tutti gli altri miei libri. Ho cercato dunque di riassumere del materiale sufficiente a darvi gli strumenti per poter affrontare i capitoli successivi. Abbiamo già abbastanza materiale per osservare innanzitutto che il tentativo, ampiamente diffuso nei vari scritti sulla ragione pratica, di trovare un elemento analogo che stia all’azione intenzionale così come la verità sta alla credenza, si è rivelato un tentativo senza speranza fin dall’inizio. La credenza è uno stato intenzionale con delle condizioni di soddisfazione. Se queste condizioni sono soddisfatte si dice che la credenza è vera. Le credenze hanno una direzione di adattamento mente-a-mondo. Ma l’azione intenzionale consiste di due componenti: un’intenzione-in-azione e un movimento corporeo. Le azioni in quanto tali non hanno condizioni di soddisfazione. Piuttosto, ogni intenzione-in-azione ha una condizione di soddisfazione e, se soddisfatta, causerà il movimento corporeo o un altro fenomeno che costituisce il resto dell’azione. Così l’azione avrà successo se e solo se l’intenzione-in-azione è soddisfatta. Ma in aggiunta a questa condizione di soddisfazione, non c’è un’ulteriore condizione di soddisfazione per le azioni in quanto tali. Quando l’azione è premeditata, e cioè quando c’è un’intenzione precedente, l’occorrenza dell’azione stessa, in quanto causata dall’intenzione precedente, costituirà la condizione di soddisfazione dell’intenzione precedente. Sia l’intenzione precedente che l’intenzione-in-azione hanno direzioni di adattamento mondo-a-mente. Le azioni sono in effetti le condizioni di soddisfazione delle intenzioni precedenti, proprio come i movimenti corporei sono le condizioni di soddisfazione delle intenzioni-in-azione. Ma, come ho osservato prima, non tutte le azioni richiedono un’intenzione precedente poiché non tutte le azioni

sono premeditate. Tutte le azioni richiedono invece un'intenzione-in-azione e, infatti, possiamo definire un'azione umana come un qualsiasi evento complesso che contiene un'intenzione-in-azione come uno dei suoi componenti. Nei capitoli seguenti ci occuperemo di vedere come gli agenti razionali possono organizzare i loro contenuti intenzionali, così come le loro rappresentazioni dei fatti nel mondo, in modo da formare delle intenzioni precedenti e delle intenzioni-in-azione razionalmente motivate.

P41

3.

IL DIVARIO: IL TEMPO E IL SÉ

I. AMPLIARE IL DIVARIO

L'esistenza del divario ci lascia con un certo numero di questioni aperte. Eccone una: nello spiegare le azioni fornendo delle ragioni, normalmente non facciamo appello a condizioni causalmente sufficienti. Ma se le cose stanno in questo modo, come può allora la spiegazione veramente spiegare qualcosa? Se gli antecedenti causali sono insufficienti a determinare l'azione, come può il fare appello ad essi spiegare il perché si è verificata quest'azione piuttosto che una qualsiasi altra azione, che era ugualmente possibile, dato il medesimo insieme di cause antecedenti? La risposta a questa domanda ha delle implicazioni profonde e nel corso di questo capitolo cercherò di risolverne alcune.

Il mio primo obiettivo sarà quello di provare a stabilire, al di là di ogni ragionevole dubbio, che esiste veramente quel fenomeno del divario (*gap*) di cui abbiamo già discusso. Per fare questo devo fornire una definizione più precisa del divario e dire qualche cosa di più sulla sua geografia. Il mio secondo obiettivo sarà rispondere alla domanda che ho appena posto e abbozzare alcune implicazioni della risposta. Sosterrò che per spiegare i fenomeni del divario dobbiamo presupporre una nozione del sé che sia irriducibile, non humanea, e dobbiamo presupporre l'esistenza di certe relazioni speciali tra il sé e il tempo, per quanto riguarda la ragione pratica.

La definizione del divario

Si possono dare due descrizioni equivalenti del divario, una rivolta al futuro, l'altra rivolta al passato. Quella rivolta al futuro è la seguente: il divario è quella caratteristica del nostro prendere decisioni coscienti e del nostro agire in cui abbiamo la sensazione di disporre di decisioni future alternative e di azioni che sono causalmente aperte dinanzi a noi. Quella rivolta al passato: il divario è quella caratteristica del nostro prendere decisioni coscienti e del nostro agire tramite la quale le ragioni che precedono le decisioni e le azioni

non vengono esperite dall'agente come se costituissero delle condizioni causalmente sufficienti per le decisioni e le azioni. Per quanto riguarda le nostre esperienze coscienti, il divario ha luogo quando le credenze, i desideri e altre ragioni non vengono esperite come condizioni causalmente sufficienti per una decisione (la formazione di un'intenzione precedente); il divario ha luogo anche quando l'intenzione precedente non costituisce una condizione causalmente sufficiente per un'azione intenzionale; si presenta anche quando l'inizio di un progetto intenzionale non costituisce un insieme di condizioni sufficienti per la sua continuazione o per il suo completamento.

La geografia del divario

Queste tre manifestazioni del divario illustrano la sua geografia di base. Per prima cosa, quando una persona sta prendendo una decisione razionale, c'è un divario tra il processo deliberativo e la decisione stessa, dove la decisione consiste nella formazione dell'intenzione precedente. In secondo luogo, una volta che una persona si è decisa a fare qualcosa, vale a dire, una volta che ha formato un'intenzione precedente, esiste un divario tra l'intenzione precedente e l'effettivo cominciamento dell'azione, nell'inizio di un'intenzione-in-azione. Terzo, ogniqualvolta una persona sta svolgendo un qualche tipo di attività estesa nel tempo, così come io ora sto scrivendo un libro, c'è un divario tra le cause, nella forma di un'intenzione precedente alla realizzazione dell'azione, e l'intenzione-in-azione, da un lato, e l'effettivo portare a compimento l'attività complessa, dall'altro. Quando si parla di azioni estese nel tempo, anche una volta date le vostre intenzioni precedenti e il cominciamento dell'azione in una intenzione-in-azione, dovrete ancora continuare a provare; dovrete continuare da voi. Tutti e tre i divari possono essere visti come aspetti diversi della medesima caratteristica della coscienza, quella caratteristica per la quale le nostre esperienze coscienti della decisione e dell'azione (l'esercizio della volontà, la sensazione cosciente dello sforzo – questi sono tutti nomi per la stessa cosa) non vengono esperite come aventi delle condizioni causali psicologicamente sufficienti per far sì che esse accadano.

II. ARGOMENTI A FAVORE DELL'ESISTENZA DEL DIVARIO

Ci sono, mi sembra, tre tipi di scetticismo che si potrebbero avere riguardo al divario. Per prima cosa, potrei aver descritto in maniera imprecisa la coscienza. Forse non esiste un divario del genere. Seconda cosa, ammesso che esista, la psicologia dell'inconscio probabilmente non tiene conto dell'esperienza cosciente della libertà. Le cause psicologiche possono essere sufficienti a determinare tutte le nostre azioni, persino quando non siamo coscienti di queste cause. Terza cosa, anche se fossimo liberi da un punto di vista psicologico, questa libertà potrebbe essere un epifenomeno. La neurobiologia, che soggiace a tutto questo, potrebbe determinare tutte le nostre azioni. Dopotutto, non ci sono divari nel cervello. In questo capitolo risponderò al primo tipo di scetticismo, nel capitolo 9 risponderò al terzo. Non ho niente da dire riguardo al secondo, poiché non lo considero seriamente. Ci sono infatti alcuni casi in cui le nostre azioni sono determinate da cause psicologiche inconscie – i casi di ipnosi, per esempio – ma sembra alquanto incredibile sostenere che tutte le nostre azioni sono simili all'agire in uno stato ipnotico. Ho discusso brevemente questa questione in un altro libro¹ e in questa sede non aggiungerò altre cose.

La prova più semplice di quanto sto descrivendo come la causa speciale e come gli elementi di volizione del divario si trova nel seguente esperimento mentale, basato sulla ricerca di Wilder Penfield². Egli scoprì che stimolando la corteccia motoria dei suoi pazienti con un micro elettrodo, poteva causare dei movimenti corporei. Quando venivano interpellati, i pazienti rispondevano sempre: “non l’ho fatto io, l’hai fatto tu” (pag. 76). Così, ad esempio, l’esperienza del paziente di avere il braccio alzato, a causa della stimolazione cerebrale operata da Penfield, è alquanto differente dalla sua esperienza di alzare il braccio volontariamente. Qual è la differenza? Bene, per rispondere, immaginate i casi di Penfield su larga scala. Immaginate che tutti i miei movimenti corporei, che si svolgono in un certo periodo di tempo, siano causati da un neuroscienziato che invia dei raggi elettromagnetici nella mia corteccia motoria. Ora, chiaramente, l’esperienza sarebbe completamente differente dalla normale azione volontaria cosciente. In questo caso, come nella percezione, io osservo ciò che mi sta accadendo. Normalmente io *faccio sì che questo accada*. Ci sono due caratteristiche del caso normale. Per prima cosa, provando a sollevare il mio braccio io causo il movimento corporeo. Il tentativo è sufficiente a causare il movimento del mio

¹ *Minds, Brains, and Science*, Cambridge, MA, Harvard University Press, 1984. Si veda il capitolo 6; *Mente, cervello, intelligenza*, tr. it. Bompiani, Milano 1988.

² W. Penfield, *The Mystery of the Mind*, Princeton, Princeton University Press, 1975, pp. 76-77, tr. it *Il mistero della mente*, Vallecchi, Firenze, 1991, pp. 90-91.

braccio; ma, in secondo luogo, le ragioni per l'azione non sono cause sufficienti per far sì che il tentativo stesso sia obbligatorio.

Se proviamo ad osservare tutto questo attraverso una lente di ingrandimento, troveremo che l'azione consiste di due componenti che ho descritto nel capitolo 2, l'intenzione-in-azione (il tentativo) che, quando è cosciente, è l'esperienza cosciente dell'agire, e il movimento corporeo. L'intenzione-in-azione è causalmente sufficiente per il movimento corporeo. Così, se io sollevo il mio braccio, l'intenzione-in-azione causa il fatto che il mio braccio si sollevi. Ma in un caso normale di azione volontaria, l'intenzione-in-azione non ha essa stessa delle condizioni precedenti causalmente sufficienti dal punto di vista psicologico, e quando dico che l'intera azione è priva di condizioni sufficienti è perché è l'intenzione-in-azione a esserne priva. Questa è una manifestazione del divario della libertà umana. Nel caso normale, l'esperienza dell'agire causerà l'inizio del movimento tramite condizioni sufficienti, ma quella stessa esperienza (l'esperienza del provare ciò che William James chiamava la sensazione dello "sforzo") non avrà sufficienti condizioni causali psicologiche nei casi di libertà e di volontà.

Nel primo capitolo ho brevemente accennato ad un secondo argomento: io credo che nella vita reale le manifestazioni più drammatiche del divario si presentino quando si hanno ragioni diverse per compiere un'azione, o per scegliere un'azione, e si può invece agire in un unico modo; si può scegliere una sola ragione sulla base della quale agire. Supponete, per esempio, che io abbia diverse ragioni per votare un particolare candidato politico. Non posso tuttavia votare per lo stesso candidato per tutte queste ragioni. Posso votare per quel candidato per una sola ragione, e per nessun'altra. In questo caso, posso sapere senza osservazione che ho votato per quel candidato per una ragione particolare, e per nessun'altra, anche se sapevo di avere molte altre ragioni per votarlo. Ora, questo è un fatto sorprendente e noi dobbiamo esaminarlo. Ci sono diverse ragioni che operano in me, ma soltanto una di queste è realmente effettiva ed *io seleziono quella che sarà effettiva*. Vale a dire, per quanto riguarda la consapevolezza delle mie azioni, le mie diverse credenze e desideri non *causano* il fatto che io mi comporti in un certo modo. Piuttosto sono io che seleziono il desiderio sulla base del quale agire. In breve, io decido quale delle molte cause sarà quella effettiva. Ciò suggerisce un'ipotesi affascinante che emergerà anche nei prossimi capitoli. Se pensiamo alle ragioni sulla base delle quali agiamo come alle ragioni che sono *effettive*, allora emerge che dove è implicata l'azione razionale libera *tutte le ragioni effettive sono*

rese effettive dall'agente, dal momento che egli sceglie le ragioni sulla base delle quali agire.

Quando affermo che noi “selezioniamo” le ragioni in base alle quali agire, o che noi “rendiamo” le ragioni effettive, non intendo sostenere che l'atto del selezionare e quello del far accadere siano degli atti separati. Se così fosse, potremmo rapidamente costruire degli argomenti viziosi di regresso all'infinito sul rendere effettivo il rendere effettivo stesso e sul selezionare le selezioni stesse³. Voglio solamente dire che quando si agisce liberamente sulla base di una ragione che è stata precedentemente selezionata, con quello stesso atto si seleziona quella ragione e la si rende effettiva.

Un terzo modo, più indiretto, per sostenere l'esistenza del divario è notare che la razionalità è possibile solamente quando è possibile l'irrazionalità. Ma la possibilità di entrambi richiede la libertà. Così, posso comportarmi razionalmente solo quando sono libero di compiere un certo numero di scelte possibili e solo se ho davanti a me la possibilità di comportarmi in modo irrazionale. Paradossalmente, il presunto ideale di una macchina perfettamente razionale, il computer, non è affatto un esempio di razionalità, poiché il computer si trova completamente al di fuori dell'ambito della razionalità. Un computer non è né razionale né irrazionale, poiché il suo comportamento è interamente determinato dal suo programma e dalla struttura del suo *hardware*. L'unico senso in cui il computer può dirsi razionale è dal punto di vista relativo all'osservatore.

III. LA CAUSAZIONE E IL DIVARIO

Per indagare la relazione tra il divario e la causazione, focalizziamo la nostra attenzione sul divario compreso nella struttura reale delle azioni volontarie. Quando compiamo delle azioni volontarie coscienti, abbiamo quasi sempre la sensazione di poter disporre di possibilità alternative. Per esempio, proprio in questo momento sono seduto al computer, e sto digitando delle parole che appaiono sullo schermo. Ma potrei fare una moltitudine di altre cose. Potrei alzarmi e camminare intorno, leggere un libro o digitare delle parole che non siano queste. Supponiamo che voi stiate leggendo questo libro seduti su

³ Gilbert Ryle è noto per questo genere di argomenti del regresso all'infinito contro la tradizionale teoria dell'azione. Si veda il suo *The Concept of Mind*, New York, Harper and Row, 1949; tr.it. *Lo spirito come comportamento*, Laterza, Roma-Bari, 1982.

di una sedia. A meno che la vostra situazione non sia particolare – per esempio siete legati alla sedia o siete paralizzati – anche voi avrete l'impressione di poter fare una moltitudine di altre cose. Potreste leggere qualcos'altro, chiamare al telefono un vostro vecchio amico o uscire a prendervi una birra, giusto per accennare a poche possibilità. Questa sensazione di possibilità alternative è costruita proprio sulla struttura delle azioni umane ordinarie e ci fa avere la convinzione – o, forse, l'illusione – della libertà. Non sappiamo come sia la vita cosciente degli animali, ma la neurofisiologia degli animali più evoluti è così simile alla nostra che dobbiamo supporre che le esperienze caratteristiche dell'azione volontaria umana siano condivise da molte altre specie.

Se vivessimo la vita degli alberi o delle pietre coscienti, capaci di percepire ciò che ci sta intorno ma incapaci di una qualsiasi azione spontanea, non avremmo le esperienze che ci danno la convinzione del nostro libero arbitrio. Non tutte le esperienze, nemmeno le esperienze dei nostri stessi movimenti, contengono questo senso di libertà. Se agiamo in preda a potenti emozioni, come ad esempio una rabbia assoluta, non avvertiamo la sensazione di poter fare qualcos'altro. Peggio ancora, se le cose sono completamente al fuori dal nostro controllo, se stiamo cadendo da un edificio o se il nostro corpo è appeso immobile, non avvertiamo la sensazione di possibilità alternative, almeno non le possibilità alternative del movimento fisico.

Nella percezione, contrariamente all'azione, non abbiamo niente di simile a questo senso di possibilità alternative davanti a noi. Al contrario, diamo per scontato che le nostre esperienze percettive sono fissate dalla combinazione di come il mondo è e di come noi siamo. Per esempio, se guardo in basso i tasti del computer, non dipende da me ciò che vedo. Sebbene ci sia un elemento volontario nella percezione (per esempio, nel cambiamento di percezioni tipico della *Gestalt*, posso liberamente scegliere di vedere una figura ora come un'anatra, ora come un coniglio), in casi come questo posso dire che le mie esperienze visive sono interamente determinate da cose come la struttura della tastiera, le condizioni luminose e il mio apparato percettivo. Naturalmente, io posso sempre girare la testa, ma questa è un'azione volontaria, non un atto di percezione. Si noti il contrasto tra la libertà dell'azione e la natura determinata della percezione. La produzione delle lettere che sto inserendo, qui ed ora, nello schermo del computer dipendono da me, e posso produrne altre a volontà, dal momento che le lettere che vedo sulla tastiera sono fissate dalla fisica

della macchina. Ma che cosa significa dire che avvertiamo un senso di libertà? Quali sono le implicazioni di tale senso?

Un'altra caratteristica che pervade le nostre esperienze è l'esperienza della causazione. Nell'azione cosciente e nella percezione cosciente spesso abbiamo esperienza delle nostre relazioni con il mondo come di relazioni causali nella loro struttura essenziale. Nell'azione abbiamo esperienza di noi stessi mentre agiamo causalmente su cose che stanno al di fuori di noi, e nella percezione abbiamo esperienza di cose che stanno nel mondo e che agiscono causalmente su di noi. E qui abbiamo l'anomalia introdotta dall'esperienza dell'azione volontaria: il senso di libertà presente nell'azione volontaria è una sensazione per la quale le cause dell'azione, sebbene effettive e reali nella forma di ragioni per l'azione, sono insufficienti a determinare il fatto che l'azione accadrà. Posso dirvi perché sto facendo ciò che sto facendo ora, ma nel dirvi il perché non cerco di dare una spiegazione causalmente sufficiente del mio comportamento, perché se così fosse, la spiegazione sarebbe inevitabilmente incompleta. Potrebbe essere soltanto una spiegazione causale *parziale* del mio comportamento, perché nello specificare queste cause non vi fornisco quelle che penso siano le condizioni causalmente sufficienti. Se mi chiedete: "Perché stai scrivendo su quest'argomento?" Io vi risponderò: "Voglio spiegare alcune caratteristiche particolari dell'azione volontaria". Questa risposta che risulta una completa e adeguata spiegazione del mio comportamento, potrebbe essere solo una parte della *spiegazione causale* del mio comportamento presente, perché non specifica una causa sufficiente a *determinare* le mie azioni presenti. Persino se fornissi in ogni dettaglio le mie credenze e i miei desideri per spiegare ciò che sto facendo, anche fornendo questo insieme totale di cause, il mio comportamento non sarebbe ancora completamente determinato, e avrei ancora la sensazione di poter fare qualcos'altro. La conseguenza di ciò è che la spiegazione del nostro comportamento ha una caratteristica particolare: le spiegazioni che normalmente forniamo quando affermiamo le ragioni delle nostre azioni sono spiegazioni causali *non sufficienti*. Esse non mostrano che ciò che è accaduto doveva per forza accadere.

Come abbiamo osservato nel capitolo 1, comunemente si dice che le azioni sono causate dalle credenze e dai desideri; ma se la "causa" implica la nozione di "causalmente sufficiente", allora per quanto riguarda l'esperienza dell'azione volontaria, questo enunciato

è semplicemente falso. In *Dell'intenzionalità*,⁴ ho provato a spiegare alcuni dei sorprendenti parallelismi tra la struttura intenzionale dei fenomeni cognitivi come la credenza, la memoria e la percezione da un lato, e i fenomeni volitivi come il desiderio, l'intenzione precedente e l'azione intenzionale, dall'altro. Ho riassunto alcune di queste caratteristiche di base della struttura dell'intenzionalità nel capitolo 2 di questo libro. Ciò che il capitolo mostra è che per quanto riguarda la struttura formale dell'intenzionalità, inclusa la causazione intenzionale, la cognizione e la volizione sono immagini speculari l'una dell'altra. Queste relazioni sono illustrate nella tabella del capitolo 2. Penso che i parallelismi siano esatti, ma in questo momento vorrei richiamare l'attenzione su di una differenza: la volizione, caratteristicamente, contiene il divario in un modo diverso dalla cognizione.

IV. IL DIVARIO ESPERIENZIALE, IL DIVARIO LOGICO E IL DIVARIO INEVITABILE.

Supponiamo che fino a questo punto io sia nel giusto: c'è un divario di cui si ha esperienza ed esso è definito in relazione alla causazione intenzionale, ma l'esperienza è quella dell'assenza di condizioni causali sufficienti. Mi sembra che qualcuno potrebbe dire: "E allora? Tu hai queste esperienze, ma finora non è stata fornita alcuna ragione per cui dovrebbero interessarci o per cui esse non potrebbero essere delle illusioni sistematiche. Anche noi abbiamo delle esperienze che riguardano il colore, ma vi sono persone che pensano che la fisica abbia mostrato che il colore è un'illusione. È un'illusione che non possiamo fare a meno di avere, ma rimane comunque un'illusione. Perché il divario dovrebbe essere invece qualche cosa di diverso?"

Per quanto riguarda tutto quello che ho detto finora, il divario potrebbe essere un'illusione ma, diversamente dalla credenza in un'esistenza oggettiva dei colori, non si tratta di una credenza alla quale possiamo rinunciare. L'interesse della discussione non è di tipo semplicemente "fenomenologico". Dobbiamo presupporre che esista realmente un divario, che la fenomenologia corrisponda a una realtà ogniqualevolta ci impegniamo in una scelta e prendiamo una decisione, e che non possiamo evitare di scegliere e di decidere. Io

⁴ John R. Searle, *Dell'intenzionalità. Un saggio di filosofia della conoscenza.*, tr. it. Bompiani, Milano 1985.

posso rinunciare intelligibilmente alla mia credenza nella realtà e nell'esistenza oggettiva dei colori come qualcosa in aggiunta ai riflessi della luce, ma non posso, in questo modo, rinunciare alla mia credenza nella realtà del divario.

Sto qui proponendo tre tesi.

1. Noi abbiamo esperienze del divario del tipo che ho descritto.
2. Dobbiamo presupporre il divario. Dobbiamo presupporre che gli antecedenti psicologici di molte delle nostre decisioni e delle nostre azioni non costituiscano un insieme di condizioni causalmente sufficienti per quelle decisioni e quelle azioni.
3. Nella normale vita cosciente una persona non può evitare di scegliere e di decidere.

Per i punti 2 e 3 espongo il seguente argomento: se realmente pensassi che le credenze e i desideri sono sufficienti a causare l'azione allora potrei semplicemente sedermi e guardare l'azione svilupparsi, come quando mi siedo al cinema e guardo l'azione che si sviluppa. Ma non posso fare questo quando sono impegnato in una decisione razionale e nell'azione. Devo presupporre l'insieme precedente di condizioni psicologiche non sia causalmente sufficiente. Inoltre, ecco un argomento in più per il punto 3: anche se mi convincessi della falsità della tesi del divario, dovrei ugualmente avere a che fare con le azioni e quindi esercitare la mia libertà, indipendentemente da che essa cosa riguardi. Supponiamo che io mi convinca che non esiste il divario; dovrei ugualmente fare qualcosa, e nel fare qualcosa io esercito la mia libertà, se non altro per quanto riguarda la mia esperienza del divario. Come abbiamo visto nel capitolo 1, persino il rifiuto ad esercitare la libertà mi risulta intelligibile, in quanto agente, solo se lo considero in quanto esercizio della libertà.

Per esempio, c'è una sorta di inconsistenza pratica nel sostenere le seguenti due tesi:

1. Sto cercando di decidermi per chi votare alle prossime elezioni.
2. Ritengo che le cause psicologiche esistenti, che operano su di me proprio in questo momento, siano causalmente sufficienti a determinare per chi voterò.

L'inconsistenza viene dal fatto che se io davvero credo che (2), allora non sembra esserci un motivo per compiere lo sforzo racchiuso in (1). La situazione sarebbe analoga a quella di prendere una pillola che sono certo da sola guarirà il mio mal di testa, e poi aggiungere qualche ulteriore sforzo psicologico agli effetti della pillola. Se davvero credo che la pillola sia sufficiente allora la cosa razionale da fare è sedermi e lasciare che essa faccia effetto.

Supponiamo che io creda nella dottrina secondo la quale le azioni razionali sono causate da credenze e desideri. Supponiamo, ipotesi fantascientifica, che esistano delle pillole che inducono credenze e desideri. Ora, supponiamo che io voglia che qualcuno faccia qualcosa razionalmente. Voglio che lui voti per il candidato democratico per una ragione particolare, e così gli do le pillole rosse, che lo portano ad avere un desiderio di votare per il candidato che egli ritiene il migliore per l'economia, e gli do le pillole blu che lo convincono che il candidato democratico è il migliore per l'economia.

Ora posso semplicemente sedermi e osservare in che modo lavorano le cause? Sarebbe proprio come mettere della dinamite sotto un ponte, accendere la miccia e guardare il ponte esplodere? No. Anche in questo caso non è così, poiché supponiamo che io voglia indurre me stesso a votare per i Democratici e che quindi prenda sia le pillole rosse che quelle blu. Dopo un paio di settimane potrei pensare che le pillole abbiano funzionato bene. Sono giunto a credere che il candidato democratico è il migliore per l'economia e sono arrivato a desiderare un candidato che sia il migliore per l'economia. Ma questo non è ancora sufficiente. Devo ancora decidere per chi ho intenzione di votare e ciò presuppone che le cause non siano ancora sufficienti.

Per riassumere questi punti: noi abbiamo l'esperienza della libertà, dobbiamo presupporre l'esistenza della libertà ogniqualvolta prendiamo delle decisioni e compiamo delle azioni, e non possiamo evitare di prendere decisioni e di compiere delle azioni.

V. DAL DIVARIO AL SÉ

Nel caso dell'azione volontaria, le cause psicologiche non rendono necessario l'effetto. Allora che cosa lo rende necessario? A livello psicologico, nulla. L'effetto non è necessario, è volontario. Ciò che rende un'azione psicologicamente libera è precisamente il

fatto che le cause psicologiche precedenti non erano sufficienti a causarla. Forse, a qualche differente livello descrittivo, forse al livello delle sinapsi e dei neurotrasmettitori, le cause sarebbero sufficienti per i movimenti corporei, ma al livello di descrizione dell'azione intenzionale, la definizione di azione libera (volontaria, razionale, cosciente) è che essa non ha antecedenti psicologici causalmente sufficienti. L'errore è pensare che noi dobbiamo trovare qualcosa che renda necessario l'effetto. Ciò è sbagliato. L'effetto è l'intenzione-in-azione cosciente e cioè l'esperienza dell'agire.

Ma che cosa significa dire che l'effetto è volontario e non necessario? Che cosa potrebbe significare? Negli esempi che abbiamo considerato, questo è dato dal fatto che io elabori la decisione di fare qualcosa e che faccia poi quel qualcosa. Le ragioni per l'azione non sono causalmente sufficienti ed io agisco in base al presupposto che esse non siano causalmente sufficienti. Come dobbiamo descrivere, allora, ciò che accade? In che modo l'azione ha luogo se niente colma il divario? *L'intelligibilità della nostra operazione all'interno del divario richiede una nozione irriducibile del sé.*

Questa è una tesi molto importante per lo sviluppo di questo libro e voglio quindi cercare di chiarirla e di giustificarla. Cominciamo dunque ribadendo il contrasto con la percezione. Quando vedo qualcosa, io non devo effettivamente fare nulla. Assumendo che il mio apparato percettivo sia intatto e che io mi trovi in una posizione adeguata, io semplicemente ho delle esperienze percettive. La mia sequenza di esperienze ne include una che prima non c'era. Ma questo è tutto ciò che accade. Ora supponiamo che io stia cercando di decidere che cosa fare. Io non posso semplicemente aspettare e osservare ciò che accade. Devo realmente fare qualcosa, anche se questo è semplicemente prendere una decisione. Quando apro il mio armadio per vedere se la mia camicia è là, non devo fare altro che guardare; il resto viene da sé. Ma per indossare la camicia io devo compiere effettivamente uno sforzo. Devo avere un'intenzione-in-azione. L'intelligibilità di questo processo, assieme al suo risultato, richiede di postulare un'entità che non è richiesta per la percezione. Perché? Ebbene, sono io stesso che devo compiere questo processo, esso non accadrà di per sé.

Dobbiamo distinguere:

1. L'azione, semplicemente, accade
2. Io compio l'azione

(1) non è una descrizione corretta dell'azione umana volontaria. Un'azione del genere non accade semplicemente. Piuttosto, (2) è corretta: devo compiere l'azione affinché essa accada. Ma (2) non è un argomento causale? Per ogni argomento causale, si chiede sempre "che cosa, esattamente, causa qualcos'altro?" E in questo caso non c'è una risposta a questa domanda. Vi erano delle mie caratteristiche personali che, insieme alle mie credenze e ai miei desideri, erano sufficienti a causare l'azione? Forse sì, ma se le cose stanno così questa non è una parte dell'esperienza dell'agire, poiché io non posso sedermi e lasciare che una mia caratteristica personale svolga il lavoro. Io devo, come si dice, prendere una decisione e quindi compiere l'azione. Il fatto che io prenda la decisione e compia l'azione non significa che c'era qualche evento in me che, assieme alle mie ragioni, era causalmente sufficiente per la decisione e per l'azione.

VI. LA DESCRIZIONE SCETTICA DEL SÉ DI HUME.

Passerò ora ad analizzare tali questioni in dettaglio. Con la più grande riluttanza sono arrivato alla conclusione che noi non possiamo farci un'idea del divario, del ragionamento, dell'azione umana e della razionalità in generale senza una nozione del sé irriducibile, ossia non *humeana*. Ora mi rivolgo alla questione del sé e, dal momento che l'argomento ha bisogno di essere sviluppato attentamente dirò qualcosa circa il problema tradizionale del sé nella filosofia e circa la concezione neo-*humeana*, che è più o meno accettata nella nostra tradizione filosofica e, fino a poco tempo fa, persino da me.

Il sé è una delle nozioni più scandalose in filosofia. Non c'è niente di sbagliato nella nozione del sé come appare nel discorso ordinario. Quando diciamo, per esempio, "mi sono tagliato" o "compatire sé stessi è un vizio" il concetto di "sé" è semplicemente un'abbreviazione per i pronomi personali appropriati e per altre espressioni che si riferiscono alle persone e agli animali. Non ha alcuna pregnanza metafisica. Ma in filosofia la nozione è stata usata in relazione a diverse questioni di carattere metafisico, delle quali non tutte possono essere giustificate. Tra i concetti metafisici del sé in filosofia vi sono i seguenti:

1. Il sé è il depositario dell'identità personale nel corso del tempo. Io, al tempo t_2 , sono la stessa persona che ero al tempo t_1 poiché il sé è lo stesso. L'identità del sé descrive l'identità della persona.
2. Il sé è realmente identico all'anima. Perciò, poiché l'anima è differente dal corpo, il sé può sopravvivere alla distruzione del corpo. Il mio corpo è una cosa, la mia anima, o il sé, è qualcos'altro. Il corpo è mortale, l'anima o il sé è immortale.
3. In relazione a (1), il sé è ciò che mi rende la persona che sono. C'è una certa entità al mio interno che costituisce la mia identità come persona e mi distingue da tutte le altre persone, e questa è il mio sé. Secondo questa concezione, il sé è costitutivo del mio carattere e della mia personalità.
4. Il sé è il depositario di tutte le mie proprietà mentali. In aggiunta ai miei pensieri, ai miei sentimenti, ecc. c'è un sé che ha tutti questi pensieri e sentimenti.

Senza dubbio vi sono molti altri compiti che il sé può svolgere. Ma molti filosofi, ed io sono uno di questi, non potrebbero mai trovare una ragione sufficiente per postulare l'esistenza del sé come qualcosa che si aggiunge alla sequenza delle esperienze e al corpo in cui esse accadono. Questo tipo di scetticismo riguardo al sé è ispirato da Hume. Come egli ha posto in evidenza, quando rivolgo la mia attenzione all'interno, trovo pensieri e sentimenti particolari, ma nulla che si possa aggiungere a questi e che si definisca come il sé. Il sé, secondo Hume, è semplicemente un fascio di esperienze e niente di più. La tesi di Hume, credo, non è semplicemente che di fatto io non trovo un sé quando rivolgo l'attenzione all'interno ma, piuttosto, che niente potrebbe valere come esperienza del sé, perché ogni esperienza che avrei sarebbe semplicemente un'altra esperienza. Supponiamo che io abbia avuto un'esperienza costante che accompagnava tutte le altre mie esperienze. Supponiamo che io abbia avuto un'esperienza continua della presenza di un punto giallo all'interno del mio campo visivo. Supponiamo che essa sia durata per tutta la mia vita. Essa è il sé? No, è semplicemente un punto giallo. Non solo non c'è un'esperienza del sé, ma non potrebbe esserci una tale esperienza, poiché niente potrebbe logicamente soddisfare i limiti insiti nella nozione metafisica del sé.

La descrizione del sé fornita da Hume secondo la quale esso è semplicemente un fascio di percezioni necessita di una revisione sotto almeno un aspetto, per spiegare un'obiezione sollevata da Kant. Tutte le mie esperienze giungono a me stesso, in ogni

singolo momento, come parte di un campo di coscienza unificato. La mia vita cosciente ha quella che Kant, con il suo dono particolare per le frasi folgoranti, chiamava “l’unità trascendentale dell’appercezione”. E credo che egli volesse dire questo: non ho semplicemente la sensazione della camicia sulla mia schiena e il sapore della birra sulla mia bocca, ma ho entrambi come parte di un singolo campo di coscienza unificato. Hume pensava che ogni percezione fosse separata e distinta, ma questo non può essere corretto; perché in tal caso non potremmo distinguere tra una coscienza che ha dieci esperienze – la sensazione della camicia, il sapore della birra, la vista del cielo, ecc. – e dieci differenti coscienze ognuna con una singola esperienza. Così noi dobbiamo insistere che in ogni dato punto del tempo, tutte le esperienze di una persona sono unite in un singolo campo di coscienza. Ma il campo cosciente non ci dà un sé in aggiunta al campo cosciente. C’è semplicemente il campo di coscienza unificato, in continuo sviluppo, che scorre attraverso il tempo, e ogni segmento di tempo del campo cosciente è l’unità di tutte le sue differenti componenti. Alcuni di questi stati coscienti all’interno del campo di coscienza saranno ricordi degli eventi precedenti nella storia esistenziale della sequenza degli stati coscienti. Alcuni saranno persino dei sentimenti che, nel mio caso, descriverebbe come il sentirmi me stesso. Ma ancora non possiamo trovare un sé in aggiunta alla sequenza delle esperienze.

Vorrei aggiungere a questa concezione rivisitata del sé humeano la tesi secondo la quale per avere la sequenza delle esperienze coscienti è essenziale il corpo. A questo stadio non dobbiamo preoccuparci se la richiesta che proviene dal corpo è una richiesta empirica o una questione logica. Il punto è semplicemente che la sequenza degli stati coscienti deve avere una qualche realizzazione fisica. Persino se sono un cervello in una vasca, deve esserci un cervello fisico, ad un livello minimale, e se io ho delle esperienze del mondo, allora il mio cervello deve avere qualche tipo di interazione causale con il mondo.

La spiegazione aggiornata del sé humeano è dunque la seguente: io sono un cervello all’interno di un corpo in contatto causale con il mondo. Questo cervello è capace di causare e di sostenere campi di coscienza unificati e questi stati all’interno dei campi includeranno la memoria delle esperienze coscienti precedenti. È vero che c’è qualcosa che coincide con il sentirmi me stesso, ma questa è semplicemente una sensazione come un’altra e non ha alcuna rilevanza metafisica. L’esistenza di tali sensazioni non garantisce di per sé nessuna identità attraverso il tempo, e, per quanto ne so, potrebbe esserci un gran numero di altre persone che hanno delle sensazioni identiche alla mia sensazione di che cosa si prova a

essere me stesso. Per riassumere, il “sé” è interamente riducibile a elementi più semplici. Consiste di sensazioni coscienti, compresi i ricordi e il senso di “me stesso” (senza dubbio include anche molte false credenze circa il sé). Queste sono causate da e realizzate in un sistema fisico che ha un'esistenza continua, cioè il mio cervello. Sulla base della visione neo-humeana, *semplicemente non esiste una cosa come il sé, che si aggiunga a tutto questo*. Con questo finisce la storia del sé.

VII. UN ARGOMENTO A FAVORE DELL'ESISTENZA DI UN SÉ NON HUMEANO, IRRIDUCIBILE.

Per il momento lasciamo da parte tutte le nostre considerazioni di tipo humeano e riflettiamo su come gli esseri umani prendono delle decisioni e compiono azioni nel divario. Supponiamo che io mi trovi ad una riunione e che il presidente dica “tutti quelli a favore della mozione alzino il braccio destro”. Io alzo il mio braccio. Compio l'azione di votare per la mozione alzando il mio braccio destro. Ora, che cosa ha causato la mia azione di sollevare il braccio destro? Posso fornire una spiegazione causale parziale fornendo delle ragioni per la mia azione. Volevo votare per la mozione perché ero favorevole ad essa e credevo che sollevando il mio braccio destro avrei votato per essa. In quel contesto, sollevare il braccio significava votare per quella mozione.

Finora tutto bene ma, come abbiamo osservato più volte, le ragioni non costituiscono delle condizioni causalmente sufficienti. Quindi, in che modo superiamo il divario tra le mie ragioni, intese come cause psicologiche e l'effettiva realizzazione dell'azione? Ecco le due possibilità che ho menzionato in precedenza, descritte in maniera più completa:

1. L'azione non può avere una spiegazione adeguata, di nessun tipo. L'azione semplicemente accade. Essa non ha delle cause psicologiche precedenti causalmente sufficienti, quindi, come evento psicologico, l'azione era soltanto un accadimento arbitrario o casuale.
2. L'azione ha una spiegazione psicologica adeguata anche se manca di condizioni psicologiche precedenti causalmente sufficienti. Io *ho compiuto* l'azione per una

ragione. L'ho compiuta per una ragione, anche se la ragione non fissa una causa precedente sufficiente.

La tesi (1) non può essere corretta. L'azione non era un evento casuale o arbitrario che è semplicemente accaduto all'improvviso. Infatti, il rischio di (1) è quello che molti compatibilisti, compreso Hume, usano come un argomento per il determinismo. A meno che l'azione non fosse già determinata, si dice, deve esserci stato un accadimento casuale o arbitrario, qualcosa di cui io non sono in alcun modo responsabile. Ma l'azione non era né arbitraria né determinata. Abbiamo già esposto le ragioni per rifiutare il determinismo psicologico. Dobbiamo rifiutare anche la sua apparente alternativa, la casualità e l'arbitrarietà.

E così (2) deve essere corretta. Ma che cosa significa? Ci sono, in realtà, due domande. Per prima cosa, assumendo che la tesi del divario sia corretta, che cosa significa dire che io, una persona, ho compiuto un'azione per una ragione? Qual è la forma logica della tesi secondo la quale *S* ha compiuto un'azione *A* per una ragione *R*? Per porre la domanda in modo tradizionale: quale fatto corrisponde alla tesi secondo la quale *S* ha compiuto l'azione *A* a causa di *R*? E, in secondo luogo, come può una tesi di questo tipo, che specifica la ragione con cui ho compiuto l'azione, essere mai una spiegazione adeguata se la ragione non determina l'azione? Che tipo di spiegazione è se c'è un grande buco al suo interno? Sembra che una risposta adeguata alla prima domanda possa fornire una risposta alla seconda.

Gran parte della mia discussione del Modello Classico si concentra precisamente su questo punto. Secondo il Modello Classico non può esserci il divario. La spiegazione dell'azione ci richiede di operare, in modo quantitativo, soltanto sugli eventi e di esprimere le relazioni causali che vi sono tra questi: l'evento dell'azione *A* è stato causato dagli eventi *C* e *D*, ovvero dalle credenze e desideri dell'agente. (A proposito, il fatto che le credenze e i desideri non siano degli eventi costituisce un imbarazzo sul quale spesso si glissa dicendo che le cause reali sono gli *inizi* delle credenze e dei desideri, o gli eventi che hanno causato le credenze e i desideri)⁵. Molti filosofi, pur rifiutando diversi aspetti del Modello Classico, ne rimangono ancora soggiogati proprio riguardo a questo problema. Così, Thomas Nagel,

⁵ Donald Davidson, "Actions, reasons, and causes", ristampato in *Essays on Actions and Events*, New York, Oxford University Press, 1980, pp. 3-19. "Azioni, ragioni, cause", in *Azioni ed eventi*, tr. it. Il Mulino, Bologna, 1992, 2000, pp. 41-61.

uno dei più potenti critici del Modello Classico, sostiene che se noi accettassimo il divario l'assenza delle condizioni causalmente sufficienti nella determinazione di un'azione ci spingerebbe a concludere che nell'azione libera c'è un elemento di casualità, e le nostre spiegazioni fallirebbero perché incapaci di fornire delle condizioni sufficienti. Come sostiene Nagel, una spiegazione del genere “non può spiegare precisamente ciò che si suppone spieghi, cioè *perché ho fatto quello che ho fatto piuttosto che l'alternativa che era causalmente a mia disposizione*”.⁶ Una risposta a queste domande è stata proposta da parecchi filosofi,⁷ ma essa è completamente sbagliata. La risposta è la seguente: la causa dell'azione sono *io*. Io, la persona che compie l'azione, sono la sua causa. Così non c'è alcun un divario causale. La persona, l'agente, è la causa. In alcune situazioni siamo portati a pensare a tale causazione personale (la “causa agente”, la “causa immanente”) come ad un tipo molto speciale di causazione. Secondo la descrizione di Chisholm, abbiamo bisogno di distinguere la causa agente, che egli chiama “causa immanente”, dalla causazione ordinaria dell'evento, la “causazione transeunte”. Secondo altre spiegazioni, si deve semplicemente pensare che la persona sia una causa come le altre. Ma in entrambe le versioni il divario causale è colmato da una persona che agisce in senso causale.

Credo che questa risposta sia peggiore di una filosofia errata: essa costituisce un pessimo uso del linguaggio. Sostenere che ogni volta che un qualche oggetto *x* viene citato come causa, ci deve essere una caratteristica o una proprietà di *x* o qualche evento che implica *x*, che funziona causalmente, significa porre un forte limite alla nozione di causazione. Non ha senso affermare, *tout court*, che l'oggetto *x* ha causato quel tale evento. Così, per esempio, se io dico “Bill ha causato l'incendio”, questa è un'abbreviazione per dire qualcosa come “l'accensione del fiammifero da parte di Bill ha causato l'incendio” o “la noncuranza di Bill ha causato l'incendio”. L'enunciato originale, “Bill ha causato l'incendio”, è intelligibile solo se lo considero come soggetto di un tale completamento. Ma quale dovrebbe essere il completamento nel caso di “ho causato la mia azione di sollevare il braccio”? Notate che ha perfettamente senso dire “ho causato il suo andare verso l'alto”, in

⁶ Thomas Nagel, *The View from Nowhere*, New York, Oxford University Press, 1986, pp. 116-117. *Uno sguardo da nessun luogo*, tr. it. Il Saggiatore Milano 1988, pp. 144-145.

Preoccupazioni simili sono state espresse da Galen Strawson, “Libertarianism, Action, and Self-Determination”, ristampato in T.O'Connor (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, New York, Oxford University Press, 1995, pp. 13-32.

⁷ Per esempio, Christine Korsgaard, *The Sources of Normativity*, Cambridge, Cambridge University Press, 1996, e Roderick Chisholm, *Readings in Philosophy*, Oxford, Oxford University Press, 1982, pp. 24-35. Credo che Chisholm abbia in seguito abbandonato questa prospettiva.

risposta a “che cosa ha causato il fatto che il tuo braccio sia andato verso l’alto?”. Perché in tal caso intendiamo questo come un’abbreviazione di “ho causato il suo andare verso l’alto *sollevandolo*”. In tal caso è la mia intenzione-in-azione che funziona causalmente per far sì che il mio braccio si sollevi. Ha anche perfettamente senso dire “il desiderio di votare per la mozione ha causato il sollevamento del mio braccio”. Ma questo è semplicemente per citare una ragione e ci lascia con lo stesso divario che abbiamo tentato di colmare, senza successo.

Qual è quindi la corretta interpretazione di (2)? *Il primo passo per capire (2) è notare che per la sua comprensione si richiede una nozione molto speciale di agente.* Il fascio humaneo, perfino se unificato e incarnato in un corpo, non è sufficiente. E' necessario avere un agente animale. Qualcosa è un agente in questo senso se e solo se è un’entità cosciente che ha la capacità di iniziare e di compiere delle azioni sotto il presupposto della libertà. Questo potrebbe sembrare banale, ma non è affatto ovvio e innocuo, poiché implica l’idea che per parlare di azione e di agente il fascio humaneo di impressioni e sensazioni non è sufficiente. Un agente è molto più di un tale fascio. Secondo la concezione humanea, il fascio di percezioni e sensazioni è semplicemente una sequenza di fenomeni naturali, che è parte della sequenza delle cause efficienti e degli effetti del mondo. Ma un agente in questo senso richiede molto di più che essere un fascio o l’essere parte di un fascio. Perché? Perché l’intenzione-in-azione non è semplicemente un evento che accade di per sé. Essa può accadere soltanto se un agente sta effettivamente facendo qualcosa, o almeno sta provando a fare qualcosa. *La nozione di agente richiede un’entità capace di tentare di fare qualcosa in modo cosciente.*

Ma, fino ad ora, non abbiamo ancora spiegato come o perché possiamo, o dovremmo, accettare delle spiegazioni causali non sufficienti. Passiamo dunque al passaggio successivo. Poiché l’agente deve essere capace di prendere decisioni e di compiere delle azioni in base alle ragioni, la medesima entità che agisce come agente deve essere capace di percezione, di credenza, di desiderio, di memoria e di ragionamento. Nel linguaggio filosofico tradizionale la nozione di agente era stata introdotta per descrivere la *volizione*, ma la stessa entità che ha volizione deve anche avere *conazione* e *cognizione*. L’agente deve, in breve, essere un sé. Proprio come l’agente deve essere aggiunto al fascio di percezioni e di sensazioni per spiegare come i fasci inseriti in un corpo possano impegnarsi in azioni libere, così il sé deve essere aggiunto all’agente per spiegare come gli agenti possono agire razionalmente.

La ragione per cui in questi casi possiamo accettare razionalmente delle spiegazioni che non forniscono condizioni sufficienti è che comprendiamo che le spiegazioni riguardano dei sé razionali nelle loro capacità di agenti. Così i tre seguenti enunciati appaiono simili nella loro sintassi superficiale, ma la loro semantica sottostante, come il fatto che noi li comprendiamo grazie alle nostre presupposizioni dello Sfondo, rivela importanti differenze.

- (1) Ho sollevato il mio braccio perché volevo votare a favore della mozione.
- (2) Ho mal di stomaco perché volevo votare a favore della mozione.
- (3) L'edificio è crollato perché il terremoto ha danneggiato le fondamenta.

(1) E' perfettamente accettabile come spiegazione sebbene non fornisca condizioni sufficienti, poiché noi la comprendiamo sulla presupposizione dello Sfondo dell'esistenza di sé razionali, che agiscono in base a ragioni, sotto il presupposto della libertà. Per capire questo punto mettiamo in contrasto (1) e (2). Date le nostre presupposizioni di Sfondo, (2) è interpretata come (3). Funziona come spiegazione perché, all'interno del contesto, essa fornisce delle condizioni causalmente sufficienti e la razionalità e la libertà non rientrano per nulla nello scenario. Avere mal di stomaco è una questione che non riguarda l'agire in base a una ragione.

Ma perché dovremmo accettare delle spiegazioni della forma di (1) se non forniscono condizioni causalmente sufficienti? Se c'è un divario nella spiegazione allora sembra che l'evento abbia un qualche elemento di casualità. Non è stata fornita nessuna ragione del perché è accaduto quel particolare evento in contrapposizione a qualche altro evento. Come rispondiamo all'obiezione di Nagel? La chiave della risposta è notare che la domanda "perché lo hai fatto?" richiede una risposta totalmente differente dalla domanda "perché questo è accaduto?". Ora voglio spiegarvi la differenza. Il primo passo è questo: considerate sempre il comportamento razionale e la sua spiegazione dal punto di vista della prima persona, poiché essi hanno un'ontologia in prima persona. Essi esistono solo dal punto di vista della prima persona. E da quel punto di vista non vi è alcuna possibilità per la quale si dia il caso che le ragioni non siano causalmente determinanti, ed insieme che la spiegazione sia tuttavia perfettamente adeguata di per sé. Il punto di vista della prima persona spiega sia perché io ho fatto ciò che ho fatto, sia perché ho fatto quella cosa piuttosto che un'altra che era causalmente disponibile. La prospettiva della prima persona è adeguata perché fornisce

la ragione che io, in quanto sé razionale, avevo reso effettiva agendo in base ad essa. Fornisce una risposta perfettamente adeguata alla domanda “perché lo hai fatto?” senza implicare “è causalmente impossibile che qualsiasi altra cosa sarebbe potuta accadere”. Fornisce una risposta adeguata alla domanda perché risponde precisamente alla domanda “perché?” e a “perché hai fatto questo e non qualcos’altro?” E non è un requisito necessario di questo tipo di risposta fornire delle condizioni causali determinanti. *Il divario causale non implica un divario esplicativo*. La domanda “perché lo hai fatto?” non chiede “quali cause erano sufficienti a determinare la tua azione?”, ma piuttosto, chiede: “sulla base di quale ragione (ragioni) tu, in quanto soggetto razionale, hai agito?” E la risposta a questa domanda fornisce una spiegazione non mostrando come l’azione, in quanto evento naturale, fosse inevitabile date le cause precedenti, ma *mostrando come un sé razionale ha operato all’interno del divario*. Per usare il linguaggio di Wittgenstein si potrebbe dire: questo è il modo in cui si svolge il gioco linguistico dello spiegare le azioni, e non si deve supporre che esso debba svolgersi seguendo le regole del gioco linguistico delle spiegazioni della meccanica classica. La ragione per cui il gioco linguistico dello spiegare le azioni fornendo delle ragioni si esprime in modi differenti è che i fatti reali, registrati dagli enunciati espressi in questo gioco linguistico, hanno una forma logica diversa da quella degli enunciati causali ordinari.

L’ipotesi di Nagel, nella sua formulazione è in effetti ambigua. L’ipotesi con la quale io spiego perché ho compiuto quell’azione piuttosto che un’altra che mi si presentava dinanzi potrebbe significare (a) che io formulo la ragione in base alla quale ho agito; in tal caso io formulo una ragione che spiega questa azione ed esclude altre ragioni che erano causalmente aperte dinanzi a me. Oppure potrebbe significare (b) che io esprimo le cause di un evento, la mia azione, che spiegano perché quell’evento doveva accadere e perché nessun altro evento avrebbe potuto accadere. L’obiezione di Nagel solleva un problema soltanto se supponiamo che la richiesta (b) debba essere soddisfatta se deve esserci una spiegazione. Ma questo sarebbe un errore. La domanda “perché lo hai fatto?” mi chiede di esprimere, in senso rilevante, *la ragione(ragioni) in base alla quale ho agito*.

Naturalmente, come ha posto in evidenza Nagel, fornire una ragione non risponde di per sé al perché ho agito in base a quella ragione e non a qualche altra ragione che avevo a disposizione. Ma questa è una diversa questione. “Perché lo hai fatto?” richiede inizialmente la ragione(i) in base alla quale ho agito. Si potrebbe sempre proseguire con le domande.

“Perché quella ragione secondo te era adeguata?” E tali domande riveleranno maggiori divari, ma la spiegazione deve a un certo punto giungere ad un fine. E questo non dimostra l’inadeguatezza della mia risposta alla prima domanda, che ammette ulteriori domande.

La richiesta di formulare le ragioni in base alle quali ho agito richiede il riferimento ad un sé. Le condizioni di verità delle frasi della forma “X ha compiuto l’azione A per la ragione R” non richiedono semplicemente l’esistenza degli eventi, degli stati psicologici e delle relazioni causali tra loro, essi richiedono un sé (che è qualcosa di più di un agente) che renda effettiva una ragione agendo su di essa. Parecchi filosofi, il più significativo dei quali è forse Korsgaard, hanno sostenuto che nelle azioni volontarie noi *creiamo* noi stessi. Se così fosse, si tratterebbe di una nozione del sé totalmente differente da quella che sto esponendo ora. Essi devono voler dire che noi creiamo il nostro carattere e la nostra personalità. Il punto che vorrei qui sottolineare non è il fatto che l’azione crea un sé, ma che l’azione *presuppone* un sé.

Secondo il Modello Classico, la spiegazione dell’azione richiede soltanto una quantificazione degli eventi. Così la forma logica di “S ha fatto A a causa della sua credenza e del suo desiderio” si esprime nel modo che segue:

C’è un qualche x tale che x è il fare A per mezzo di S e c’è un qualche y tale che y è una credenza e c’è un qualche z tale che z è un desiderio e che (gli inizi di) y e z causino x .

Il riferimento apparente ad un sé è solamente un mezzo per identificare un dato evento.

Secondo la spiegazione che sto proponendo, la forma logica di “S ha compiuto A a causa della ragione R” è in effetti ciò che appare come:

C’è un x tale che $x =$ il sé S, e c’è un y tale che $y =$ l’azione data A, e c’è un qualche z tale che $z =$ la ragione R, e x ha compiuto y e nel compiere y , x ha agito sulla base di z .

Notate che il riferimento al sé è ineliminabile. Non ho ancora spiegato che cos’è una “ragione per l’azione” e che cosa significa agire sulla base di una ragione. Questa spiegazione arriverà nel prossimo capitolo. Stiamo facendo un passo per volta, e in questo capitolo sto semplicemente cercando di rendere chiaro che la forma delle spiegazioni

dell'azione razionale non è quella della causazione tra eventi, ma richiede una nozione irriducibile del sé.

Che cosa dimostra che la mia analisi è migliore di quella del Modello Classico? Esiste un certo numero di argomenti, ma quello che stiamo considerando a questo punto ha le due seguenti premesse. Concedetemi che:

1. Le spiegazioni della ragione normalmente non forniscono condizioni causalmente sufficienti.
2. In casi normali esse sono perfettamente adeguate così come sono.

Noi sappiamo che (2) è vera in base alla considerazione degli esempi che riguardano la prima persona. Posso dirvi esattamente perché ho votato per Clinton, anche se le ragioni che fornisco non sono quelle che mi avevano obbligato a farlo. Per spiegare (2), dato (1), dobbiamo introdurre la nozione di agire in base a una ragione. La caratteristica speciale delle spiegazioni di ragione è questa:

3. La richiesta di una spiegazione di ragione dell'azione è la richiesta di un enunciato di ragione ragione sulla base del quale l'agente ha agito.

Sulla base di (3) possiamo concludere (4):

4. Tali spiegazioni richiedono la nozione di agente capace di agire sulla base di una ragione, e un qualsiasi agente di questo tipo è un sé, inteso nel senso che sto cercando di chiarire.

Il fatto che noi siamo inclini a supporre che tutte le spiegazioni debbano adattarsi ad un modello pre-costituito della causazione tipo quello delle palle da biliardo è una limitazione della nostra sensibilità di Sfondo, limitazione che sto cercando di superare. Sto cercando di spiegare le condizioni della forma particolare di intelligibilità di questo gioco linguistico.

Passiamo ora al passo successivo dell'argomentazione.

Solo grazie al sé qualcosa può essere una ragione per l'azione Finora abbiamo identificato un divario dell'esperienza e un sé che opera all'interno di quel divario. Ma il sé opera in quel divario sulla base di ragioni. Così sorge la domanda: che cos'è una ragione e quale fatto trasforma qualcosa in una ragione? Dirò molto di più sulle ragioni nei prossimi due capitoli, ma a questo punto è chiaro che affinché qualcosa sia una ragione che può funzionare nella deliberazione e nell'azione, essa deve essere una ragione per un agente. Il punto deve essere affermato in maniera precisa. Ci sono molte ragioni per fare delle cose di cui nessuno sa niente. Per esempio, la gente aveva una ragione per mangiare del pane di frumento integrale – che previene il beriberi – senza sapere che aveva una tale ragione. Ma una tale ragione non può avere un ruolo nella deliberazione. Nella deliberazione, per funzionare come una ragione, una ragione deve essere posseduta da un agente. Questa è una caratteristica aggiuntiva del sé, così come è un argomento per l'esistenza del sé. Inoltre, poiché le ragioni possono essere cognitive – credenze e percezioni, per esempio – il sé deve implicare qualcosa di più della capacità di agire, qualcosa di più di una semplice volizione. La stessa entità deve essere capace di operare sia con delle ragioni cognitive che con decisioni e azioni che si basano su quelle ragioni.

Considerato tutto questo, possiamo ora compiere il passo successivo. Se assumiamo l'esistenza di un sé cosciente irriducibile, che agisce sulla base di ragioni sotto i vincoli della razionalità e sul presupposto della libertà, possiamo ora dare senso alla nozione di *responsabilità* e a tutte le nozioni con essa concomitanti. *Poiché il sé, per prendere delle decisioni e per compiere delle azioni, opera all'interno del divario sulla base delle ragioni, esso è il luogo della responsabilità.*

Questo è un argomento separato per l'esistenza di un sé irriducibile. Affinché noi possiamo attribuire la responsabilità, deve esserci un'entità capace di assumere, esercitare e accettare la responsabilità. Comprenderemo meglio questo punto se introduciamo la nozione di tempo. La nozione di responsabilità assume un senso solo se noi possiamo attribuire ora delle responsabilità ad azioni che sono accadute nel passato. Sono ritenuto responsabile ora per delle azioni che ho compiuto in un passato remoto. Ma questo acquista senso solamente se c'è una qualche entità che può essere sia l'agente dell'azione nel passato, sia me stesso in questo momento. Questa entità è ciò che ho chiamato "il sé". Notate che io non sono responsabile delle mie percezioni in modo analogo. Le percezioni agiscono su di me ma io non sono responsabile di esse nello stesso modo in cui sono responsabile delle mie azioni.

Solamente a proposito di un sé, nel senso che abbiamo spiegato, possiamo dire che egli, o ella, è responsabile, colpevole, da biasimare, meritevole di fiducia, che si merita un premio o una punizione. Queste attribuzioni sono diverse da “è di bell’aspetto”, “è addolorato” o “guarda l’automobile che sta arrivando”. Le prime richiedono una nozione irriducibile del sé per la loro intelligibilità. Le seconde no.

Il ragionare è un processo del sé all’interno del tempo e, per la ragione pratica, il ragionare ha a che fare essenzialmente con il tempo.

L’introduzione della nozione di tempo ci permette di osservare che la razionalità in azione riguarda sempre un agente che ragiona coscientemente nel tempo, sotto il presupposto della libertà, riguardo a che cosa fare in questo momento o nel futuro. La ragione teorica, ha a che fare con che cosa accettare, concludere o credere; la ragione pratica ha a che fare con il problema di quale azione compiere. Vi è dunque un senso in cui possiamo affermare che l’intero processo del ragionare è pratico, poiché esso riguarda, in generale, il fare qualcosa. Nel caso della ragione teorica il fare è tipicamente qualche cosa che riguarda l’*accettare* una conclusione o un’ipotesi sulla base dell’argomentazione o dell’evidenza. La ragione teorica è, in questo senso, un caso speciale della ragione pratica. La differenza tra ragione teorica e ragione pratica è la direzione di adattamento della conclusione: mente-a-mondo, nel caso in cui si traccia una conclusione dall’evidenza o dalle premesse, e mondo-a-mente nel caso in cui si prende una decisione e quindi ci si forma un’intenzione sulla base di certe considerazioni. Questo comporta delle ulteriori importanti conseguenze: il ragionamento pratico non è semplicemente qualcosa che accade nel tempo, ma esso riguarda il tempo nel senso che esso è il ragionare, ora, da parte di un sé, riguardo a ciò che un sé sta per fare ora o in futuro. Così, una volta che introduciamo la nozione di tempo vediamo che il sé viene richiesto sia come il luogo della responsabilità per le azioni passate, sia come il soggetto della pianificazione di azioni presenti e future. Quando, nel presente, elaboro dei piani per il futuro, il soggetto della pianificazione è lo stesso sé che sta per compiere l’azione nel futuro. La strutturazione del tempo, che è una parte essenziale della ragione pratica, presuppone un sé.

VIII. SOMMARIO DELL'ARGOMENTO PER L'ESISTENZA DI UN SÉ IRRIDUCIBILE, NON HUMEANO.

Punto 1. L'esistenza di azioni volontarie e intenzionali, richiede un agente cosciente che agisce. Diversamente l'azione sarebbe soltanto l'accadere di un evento. Né un fascio humeano, né una "persona", nel senso di Strawson,⁸ avente proprietà sia mentali che fisiche, nemmeno una persona del tipo di Frankfurt⁹ che ha dei desideri di secondo ordine circa i suoi desideri di primo ordine, sono di per sé sufficienti a descrivere l'agente e l'azione.

Punto 2. Tuttavia è logicamente possibile essere un agente senza essere un sé. Per essere un sé l'entità che agisce come un agente deve essere anche capace di ragionamento cosciente riguardo alle sue azioni. Deve essere un'entità capace di percezione, memoria, credenza, desiderio, pensiero, inferenza, e di cognizione in generale. L'agente non è sufficiente per l'azione razionale. L'agente deve essere un sé.

Punto 3. Ed ecco il punto cruciale: c'è una caratteristica logica speciale per le spiegazioni dell'azione razionale. Se le interpretate come spiegazioni causali, esse non funzionano. Le cause non sono sufficienti per spiegare l'azione. Tuttavia sono perfettamente adeguate così come sono. La loro intelligibilità richiede il fatto che noi pensiamo ad esse non come a cause che determinano un evento, ma come a ragioni sulla base delle quali un agente razionale cosciente ha agito. Quell'agente è un sé. L'agente più l'apparato della razionalità equivale al sé.

Punto 4. Una volta che disponiamo di un sé in quanto agente dell'azione, allora potremo spiegare una quantità di altre nozioni confuse, specialmente la responsabilità, e le sue nozioni concomitanti di colpa, colpevolezza, merito, compenso, punizione, premio e condanna.

Punto 5. L'esistenza del sé spiega la relazione dell'agente con il tempo. Lo stesso sé deve essere responsabile per le azioni che ha compiuto nel passato e deve essere capace di progettare il futuro. La razionalità, in generale, accade nel tempo e la ragione pratica riguarda, nel senso che ho cercato di spiegare, il tempo.

⁸ Peter Strawson, *Individuals: An Essay in Descriptive Metaphysics*, London, Methuen, 1959, pp. 87-116. *Individui. Un saggio di metafisica descrittiva*, tr. it. Feltrinelli, Milano 1978.

⁹ Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person", *Journal of Philosophy*, gennaio 1971, pp. 5-20

IX . L'ESPERIENZA E IL SÉ

Qual é la relazione tra il sé che ho descritto, inteso come entità caratterizzabile in modo puramente formale, definita da un elenco specifico di caratteristiche, e le nostre esperienze coscienti reali? Stiamo, in un certo senso, attaccando la conclusione di Hume secondo la quale non c'è un'esperienza del sé? In breve, che cosa possiamo dire di questo "sé"? Fino ad ora, nulla. Il fatto che debba esserci un sé che agisce è un requisito formale dell'azione razionale, così come non è un requisito formale della percezione il fatto che debba esserci un agente o un sé che percepisce. Di conseguenza la descrizione humeana dell'io, o del sé, come sequenza di impressioni e di idee, per quanto rivisitata fino ad includere un corpo fisico con tutte le sue disposizioni, non coglie la caratteristica essenziale dell'agente razionale, ossia il sé.

La chiave per rispondere a questa domanda è racchiusa nell'analisi della struttura della nostra coscienza, poiché la prima condizione del sé è quella per cui esso dovrebbe essere capace di coscienza. Secondo la descrizione che sto presentando, il sé non è un'esperienza, né un oggetto di cui si ha esperienza. Quando, per esempio, osservo un tavolo, io ho un'esperienza visiva e c'è un tavolo che è l'oggetto di quell'esperienza. Diversamente, non c'è alcuna auto-esperienza né un oggetto di cui si ha esperienza in quanto sé. Piuttosto, il "sé" è semplicemente il nome per quell'entità che esperisce le sue proprie attività come qualcosa di più di un fascio inerte. È una caratteristica della mia esperienza cosciente il fatto che io mi impegni nella deliberazione e nell'azione, che io abbia delle percezioni, che usi i miei ricordi nella deliberazione, che prenda delle decisioni, che io metta in pratica le decisioni (o che non riesca a metterle in pratica) e che mi senta soddisfatto o insoddisfatto, colpevole o innocente, a seconda dei risultati di tutte queste attività. La linea di pensiero che sto seguendo è, in un certo senso, una linea sottile tra lo scetticismo di Hume e la visione pre-teoretica, ingenua secondo la quale ognuno di noi è consapevole di sé stesso come di un sé. Il punto che sto evidenziando è che sebbene il sé non sia il nome di un'esperienza né dell'oggetto di un'esperienza, tuttavia c'è una sequenza di caratteristiche formali delle nostre esperienze che sono costitutive di noi stessi in quanto sé.

Come possiamo essere sicuri che l'apparente richiesta del postulato di un sé non sia semplicemente un'illusione grammaticale attribuitaci dalla struttura soggetto-predicato delle proposizioni? Quando diciamo "io ho deciso di votare per Clinton" non stiamo reificando qualcosa per poter avere un oggetto per il termine "io" a cui riferirci? No. Perché l'esigenza grammaticale è la stessa perfino in casi in cui non sto facendo nulla. Consideriamo la frase "io vedo la rosa". Per quanto riguarda la fenomenologia, si possono descrivere fatti fenomenologici dicendo che "la sequenza di esperienze ora include quella della rosa". Ma non si cattura la caratteristica attiva della decisione dicendo che questa sequenza di esperienze ora comprende una decisione, poiché la decisione era qualcosa che io ho compiuto, un'azione intenzionale e attiva, mentre l'esperienza della rosa è soltanto passiva.

Ma non stiamo allora postulando un omuncolo che vive all'interno del divario e che prende le decisioni al nostro posto? E questo non porta forse a un regresso all'infinito? No, perché noi viviamo nei divari e prendiamo delle decisioni.

Postulare l'esistenza di un sé non richiede che noi abbiamo un'esperienza del sé. Un'analogia chiarirà questo punto. Ogniqualvolta vediamo qualcosa, abbiamo un'esperienza visiva e per spiegare l'esperienza visiva dobbiamo postulare un punto di vista dal quale l'esperienza ha luogo, anche se il punto di vista non è un'esperienza né è esso stesso esperito. Così, per esempio, per spiegare la mia esperienza visiva dell'Oceano Pacifico devo postulare che l'esperienza provenga da un particolare punto di vista nello spazio, anche se quando vedo il Pacifico non vedo il punto di vista dal quale lo osservo, né il punto di vista è parte dell'esperienza del guardare. Analogamente l'esperienza delle azioni libere richiede un sé anche se il sé non è né un'esperienza né un oggetto esperito.

Così Lichtenberg aveva torto. Non dovremmo dire "esso pensa" invece di "io penso". Se pensare è un processo volontario attivo, deve esserci un sé che pensa.

X CONCLUSIONE

Che cos'è, allora, il sé? Dal suo punto di vista, Hume aveva sicuramente ragione. Se per "sé" intendiamo un insieme di esperienze, come i dolori, o qualcosa che è oggetto delle nostre esperienze, come il tavolo di fronte a me, allora una cosa del genere non esiste. Per

spiegare l'agente razionale dobbiamo postulare un sé che combini le capacità della razionalità e dell'azione. Le caratteristiche del sé possono essere enunciate come segue:

Esiste un x tale che

1. x è cosciente.
2. x persiste attraverso il tempo.
3. x opera con delle ragioni, sotto i vincoli della razionalità.
4. x , operando con le ragioni, è capace di decidere, dare inizio e compiere azioni, sotto il presupposto della libertà.
5. x è responsabile di almeno una parte del suo comportamento.

In questo argomento è implicito un risultato che vorrei ora rendere esplicito, perché avrà una certa importanza negli ultimi capitoli. La questione della razionalità non è un problema di strutture dell'argomentazione formale, le quali risultano di un'utilità marginale. L'argomento centrale della discussione sulla teoria della razionalità è l'attività degli esseri umani (e, presumibilmente, anche di alcuni altri animali, come ci dimostrano le scimmie di Köhler), i sé, impegnati nel processo del ragionare. Proprio come il soggetto principale della filosofia del linguaggio non sono le frasi né le proposizioni, ma gli atti linguistici, così *il soggetto della filosofia della razionalità è l'attività del ragionare, un'attività diretta ad un fine svolta da dei sé coscienti.*

LA STRUTTURA LOGICA DELLE RAGIONI

Che cos'è la ragione per un'azione? Si suppone che la domanda sia terribilmente difficile, così difficile che Phillipa Foot una volta scrisse “sono sicura di non capire l'idea di una ragione per agire e mi chiedo se c'è qualcun altro che si trova nelle mie condizioni”¹. Ma perché dovrebbe essere così difficile? Dopo tutto non abbiamo a che fare con le ragioni per l'azione ogni giorno? Come può esserci un mistero? Usando lo stile di Wittgenstein si potrebbe dire: niente è nascosto.

Ebbene, niente è nascosto e, senza dubbio, la risposta è pienamente visibile. Ciononostante dobbiamo cercare di trovarla ed emergerà che la risposta è più complessa di quanto ci saremmo potuti aspettare. Dai capitoli precedenti possiamo inferire che certe caratteristiche formali sono possedute da qualsiasi entità che sia una ragione per un'azione. Per esempio l'esistenza e la realizzazione di tale entità dovrebbero essere consistenti con il divario. Essa dovrebbe essere il genere di cosa che può motivare razionalmente un'azione in modo tale che l'agente-sé possa *agire su* di essa, sebbene essa non presenti condizioni sufficienti per causare l'azione. Essa dovrebbe inoltre avere un contenuto logicamente correlato, in modo specifico, ai contenuti di un'intenzione precedente e di un'intenzione-in-azione (entrambe hanno direzione di adattamento verso l'alto) di cui era la ragione. Ma in che modo, esattamente? Tutto ciò è molto vago e penso non si possa dire nulla di realmente significativo fino a che non lavoreremo sul nostro problema più a lungo. Iniziamo dunque con il chiedere, come può qualcosa essere una ragione per qualcos'altro, e che cos'è una ragione per qualcosa? Il primo passo da fare è osservare l'uso ordinario delle proposizioni che contengono la parola “ragione” e i termini correlati come “spiegazione”, “perché?”

¹ Citato in G.Cullity e B. Gaut (edd.), *Ethics and Practical Reason*, Oxford, Oxford University Press, 1997, p. 53.

(*why?*) e “perché” (*because*). Il progetto è inizialmente quello di chiederci: sotto quali condizioni una proposizione P esprime una ragione R per un fenomeno F ? Una volta risposto a tale domanda possiamo allora andare al passo successivo, ovvero a chiederci: quali condizioni fanno sì che P esprima R per una persona che ha uno stato intenzionale come una credenza o un desiderio? E, poiché le intenzioni precedenti e le intenzioni-in-azione sono stati intenzionali, se possiamo rispondere alla domanda che riguarda gli stati intenzionali in generale, sembra allora che questo ci conduca ad una risposta riguardo ai casi speciali dell’avere intenzione di fare qualcosa. E questa risposta, se possiamo ottenerla, è già una risposta alla domanda “sotto quali condizioni P afferma una ragione R perché un agente X compia un’azione A ?”, poiché una ragione per avere intenzione di fare qualcosa o per provare a fare qualcosa è, a parità di condizioni, una ragione per farlo.

Una ragione è sempre una ragione per un agente, così sembra che stiamo cercando di completare il seguente bicondizionale:

Una proposizione P afferma una ragione R per un agente X per compiere l’azione A se e solo se...

Ma anche questa formulazione sembra troppo debole, per prima cosa perché non distingue tra le buone e le cattive ragioni, tra quelle ragioni che sono razionalmente accettabili e quelle che non lo sono; secondo, perché la nostra spiegazione delle ragioni deve distinguere, dal momento che questa formulazione non lo fa, tra quelle ragioni che sono disponibili all’agente e quelle che non lo sono. Si può avere una buona ragione per fare qualcosa senza saperlo. Per esempio, da molto tempo la gente aveva una buona ragione per non fumare le sigarette – il fumo causa il cancro – senza sapere che aveva una tale ragione. Terzo, l’uso di quella che appare come un’espressione di riferimento, “l’azione A ” è quantomeno fuorviante, perché al momento di pianificare un atto futuro, una tale azione non esiste ancora e, in realtà, potrebbe non esistere mai. Così una ragione per un’azione futura è una ragione per compiere un’azione di un certo *tipo* A . Proviamo un’altra formulazione del bicondizionale:

Un agente razionale X ritiene che una proposizione P affermi una ragione valida R perché X compia un’azione del tipo A se...

Più avanti in questo capitolo noteremo che persino questo modo di formulare la domanda è inadeguato. Come spesso accade in filosofia, il grande problema è giungere ad una corretta formulazione della domanda. Comunque, a questo punto, ci stiamo ancora girando attorno.

Notate che tali enunciati di ragione sono relazionali in tre modi. Per prima cosa, la ragione specificata è una ragione per qualcos'altro. Nulla è una ragione di per sé. Secondo, le ragioni per l'azione sono doppiamente relazionali in quanto sono ragioni per un *agente-sé* per compiere l'azione; e, terzo, per funzionare nella deliberazione, le ragioni devono essere *conosciute* all'*agente-sé*. Per riassumere, per funzionare nella deliberazione, una ragione deve essere una ragione per un certo tipo di azione, deve essere una ragione per un agente e deve essere conosciuta dall'*agente*. Tali proposizioni sono tipicamente intensionali-con-la-s perché non permettono l'inferenza secondo cui la cosa per cui la ragione è una ragione per qualcosa esiste effettivamente. Si può, ad esempio, avere una ragione per compiere un'azione che non si compirà mai (si dirà di più sull'intensionalità più avanti).

I. CHE COS'È UNA RAGIONE?

La nozione di ragione comprende almeno tre altre nozioni, e la quarta può essere compresa soltanto se considerata assieme alle altre. Le altre nozioni sono “perché?” (*why?*), “perché” o “poiché” (*because*) e “spiegazione”. Esprimere una ragione significa tipicamente fornire una spiegazione o la parte di una spiegazione. Le spiegazioni sono fornite in risposta a una domanda “Perché?” (*why?*). E una forma appropriata per fornire una ragione è “Perché (*because*)...”. Alla domanda “perché si dà il caso che *p*?” la risposta “perché si dà il caso che *q*” fornisce la ragione del perché *p*, se *q* realmente spiega, o parzialmente spiega, *p*. Questo è il motivo per cui tutte le ragioni sono ragioni “perché?”. Sia le “ragioni” che le “spiegazioni” sono nozioni di successo, nel senso che possono esserci buone o cattive ragioni/spiegazioni ma se una ragione/spiegazione originaria è davvero cattiva, essa fallisce completamente come ragione o come spiegazione.

Il “perché (*because*)” è un connettivo proposizionale non un connettivo vero-funzionale. Esso connette *inter* frasi. Anche la domanda “perché? (*why?*)” si riferisce a frasi intere. La richiesta di proposizioni intere è camuffata dal fatto che talvolta, nella grammatica superficiale della frase, la domanda “perché?” contiene una semplice espressione o una frase e la risposta “perché” contiene una frase preposizionale. Domanda:

“Perché ora?” o “Perché la barba?”. Risposta: “A causa di Sally” o “per pigrizia”. Ma in tutti questi casi noi dobbiamo intendere l'espressione più breve come abbreviazione di una frase intera. Per esempio: “perché stai partendo ora?” Risposta: “perché Sally ha bisogno di me ora”. “Perché ti stai facendo crescere la barba?” Risposta: “perché sono troppo pigro per radermi”.

La sintassi delle domande con il “perché? (why?)” e delle risposte con il “perché? (why?)”, quando pienamente scandita, richiede sempre un'intera proposizione e non semplicemente una frase sostantiva. Questa osservazione sintattica suggerisce due conseguenze semantiche. Per prima cosa, la specificazione sia dell'*explanans* che dell'*explanandum* deve avere un contenuto proposizionale intero e, secondo, deve esserci qualcosa di esterno alla frase che corrisponda a quel contenuto. Gli enunciati di ragione sono delle proposizioni, e quindi delle entità linguistiche, degli atti linguistici con particolari tipi di contenuti proposizionali; ma le ragioni stesse e gli enti dei quali sono ragioni non sono entità tipicamente linguistiche. Con alcune importanti eccezioni che menzionerò tra pochissimo, l'enunciato di ragione può fornire una spiegazione buona o adeguata solamente nel caso in cui sia l'enunciato di ragione, sia la proposizione che specifica la cosa da spiegare sono di fatto vere. Ma allora, ciò che rende l'enunciato e la proposizione veri sarà qualcosa che è indipendente dal linguaggio. Così, se mi si chiede “perché la California ha più terremoti di qualsiasi altro stato?” la mia risposta secondo la quale “la California è lo stato con il maggior numero di faglie sismiche” sarà una spiegazione solo se la California, effettivamente, risulta avere più terremoti di ogni altro stato ed essere, effettivamente, lo stato con maggiori faglie sismiche, e se queste faglie sono causalmente correlate ai terremoti. C'è un termine generale per descrivere quelle caratteristiche del mondo che rende gli enunciati e le proposizioni veri, o in virtù del quale essi sono veri, e questo termine è il “fatto”. Una spiegazione è una proposizione o un insieme di proposizioni. Ma una ragione non è una proposizione o un insieme di proposizioni, e la cosa per la quale una ragione è tale non è una proposizione o un insieme di proposizioni; piuttosto, nei casi che abbiamo considerato, sia l'*explanandum* che l'*explanans* sono dei fatti. Un fatto è una ragione solo in relazione al fatto di cui la ragione è una ragione, ed è una ragione di quel fatto solo se ha una relazione esplicativa con quel fatto.²

² I sostenitori degli atti linguistici (Dio li benedica tutti) si saranno senza dubbio chiesti perché non faccio semplicemente un'analisi dell'atto linguistico della spiegazione. Dopo tutto, spiegare qualcosa è un atto linguistico. Il motivo è che un'analisi del genere non ci

E' una tentazione, allora, pensare che *tutte* le ragioni siano dei fatti. Ma che dire dei casi dove sono in errore riguardo ai fatti, ma posso tuttavia offrire una spiegazione? Domanda: “perché hai con te un ombrello?” Risposta: “perché sta piovendo”. Sia la domanda che la risposta soddisfano la richiesta di contenuto proposizionale, ma supponiamo che io sia in errore e che non stia piovendo. Anche in questo caso c'è una spiegazione vera implicita nella mia risposta. Nel formulare l'asserto ho espresso la *credenza* che stia piovendo e che quella credenza possa costituire una ragione per la mia azione, anche se la credenza è falsa. In casi del genere possiamo affermare sia *il fatto* che la mia credenza è la ragione, sia che la *mia credenza* è la ragione della mia azione. Inoltre, io posso avere una ragione per compiere un'azione che, di fatto, non compio mai, ma se offro una ragione come spiegazione, allora essa può essere una spiegazione della mia intenzione di compiere l'azione, anche se l'azione non viene mai compiuta. Ciò che questi esempi suggeriscono è che sia le ragioni che gli enti dei quali esse sono ragioni possono essere sia dei fatti del mondo sia stati intenzionali come le credenze, i desideri e le intenzioni. Così, per esempio, la spiegazione del *perché ho affermato* che la California ha il maggior numero di faglie sismiche può essere che io *credevo* che essa avesse il maggior numero di faglie. E la mia credenza può essere una ragione per la mia azione, indipendentemente dal fatto che la credenza sia vera o falsa. Il vincolo formale di essere una ragione è che un'entità deve avere una struttura proposizionale e deve corrispondere ad un enunciato di ragione.³

L'ipotesi che questi esempi suggeriscono è la seguente: *tutte le ragioni sono entità strutturate in forma proposizionale: esse possono essere dei fatti del mondo, come il fatto che sta piovendo, o possono essere degli stati intenzionali proposizionali, come il mio desiderio di restare asciutto. Esse possono anche essere delle entità strutturate in forma proposizionale che non sono né fatti, né stati intenzionali, entità come gli obblighi, gli incarichi, le richieste e i bisogni.* Questa caratteristica dell'ontologia delle ragioni spiega il fatto sintattico che gli enunciati di ragione richiedono una clausola “che”, o una qualche altra forma equivalente, capace di esprimere un'intera proposizione. In italiano, non

darebbe risposte a quelle domande cui vogliamo rispondere in questa discussione. “Spiegare” non designa un punto illocutorio separato. Le spiegazioni sono, normalmente, degli insiemi di atti linguistici assertivi, ma per essere delle spiegazioni effettive, esse devono essere vere e i fatti che le rendono vere devono essere in una relazione esplicativa nei confronti delle cose che devono essere spiegate. Così, nessuna analisi degli atti linguistici risponderà di per sé alle domande cui vogliamo rispondere in questa sede.

³ Per un'interessante difesa della tesi secondo la quale le ragioni sono dei fatti, si veda Joseph Raz, *Practical Reason and Norms*, London, Hutchinson, 1975, cap. 1.

disponiamo di un singolo termine per nominare entità così differenti. “Fatto” e “fattivo” alludono troppo a delle verità che comprendono sia le credenze, che per qualcuno possono funzionare come ragioni anche quando sono false, che i fatti nel mondo. “Proposizione” e “entità strutturate in forma proposizionale” indicano soprattutto entità linguistiche e intenzionali. Propongo dunque di usare il vecchio termine grammaticale “fattivo” per identificare le entità che hanno una struttura proposizionale, sia che si tratti di stati intenzionali, di fatti nel mondo, o di entità che non sono né gli uni né gli altri, come gli obblighi. Stabilisco che con “entità fattiva” intendo una qualsiasi entità che abbia una struttura proposizionale, una struttura specificata da una clausola “che”. In breve, tutte le ragioni sono delle entità fattive, o, per brevità, sono dei fattivi. Così il fatto che sta piovendo, la mia credenza che sta piovendo, il mio desiderio che piovano e il mio bisogno che piovano possono essere tutte delle ragioni. Ma la pioggia in sé non può costituire una ragione. Il punto che sto mettendo in evidenza non è il punto banale per il quale tutti gli enunciati devono esprimere delle proposizioni, ma, piuttosto, che la specificazione di una ragione è essenzialmente proposizionale; e la ragione stessa, l’entità reale stessa, ha una struttura fattiva o proposizionale. Tali entità fattive includono non solo fatti del mondo come il fatto che sta piovendo ma anche credenze, desideri, bisogni, obblighi, incarichi e una moltitudine di altre entità fattive.

Così, per esempio, alla domanda: “perché stai portando un ombrello?” Io posso fornire i seguenti tipi di risposte.

1. Sta piovendo
2. Io credo che sta piovendo
3. Non voglio bagnarmi.
4. Sono obbligato a farlo.
5. Ho bisogno di restare asciutto.

Tutte queste frasi specificano delle entità fattive nel senso che ho introdotto. La prima, se vera, enuncia il fatto che sta piovendo. Ma anche la credenza, il desiderio, l’obbligo e il bisogno sono dei fattivi. Alcune ragioni rappresentano altre entità fattive. Così una credenza rappresenta un fatto nel mondo, ma la credenza può essere una ragione per qualcosa anche se non è vera, persino se il fatto corrispondente nel mondo non esiste.

Perché le ragioni devono avere una struttura fattiva? Non lo con precisione. La mia ipotesi è che si deve essere in grado di ragionare servendosi di ragioni e che si può ragionare solo con qualcosa che ha una struttura proposizionale.

La nostra prossima domanda è: che cosa rende un'entità fattiva una ragione per qualcos'altro? Posto quanto abbiamo appena affermato, la domanda equivale a chiedere: sotto quali condizioni un'entità del genere sta nella relazione di spiegazione con qualcos'altro? Da un lato c'è una classe di entità fattive, le ragioni, e dall'altro, c'è una classe di entità fattive che hanno bisogno di spiegazione e questa classe può includere fatti che riguardano quasi ogni cosa, dalle guerre ai terremoti, così come entità fattive tipo i desideri, le credenze ecc. Possiamo spiegare i membri della seconda classe enunciando alcuni membri della prima classe. Quali caratteristiche della prima classe consentono dunque loro di spiegare i membri della seconda classe? Le varietà di relazioni esplicative corrispondono alle indefinite varietà di spiegazioni dei fenomeni che è possibile fornire – causali, logiche, giustificatrici, estetiche, legali, morali, economiche ecc. Che cosa hanno in comune tutte queste, oltre alla caratteristica banale di fornire delle spiegazioni? Non lo so e forse non hanno nulla in comune. Sembrerebbe che le spiegazioni formino una famiglia, in senso wittgensteiniano, e che siano riunite da somiglianze di famiglia. C'è un numero enorme di tipi differenti di relazioni esplicative, ma c'è un elemento formale comune che le caratterizza ed è l'elemento della modalità: la famiglia modale include il perché qualcosa *doveva* essere o *doveva* accadere, o *potrebbe* o *dovrebbe* essere accaduta, o *avrebbe dovuto* essere accaduta ecc. La relazione esplicativa include il far accadere qualcosa, il causare, il rendere necessario, il rendere più probabile, il giustificare, il determinare, il fare qualcosa con il proposito di o per amore di..., e così via. Penso che la nozione più primitiva qui sia quella di far sì che qualcosa accada e che le nostre forme paradigmatiche di spiegazione siano spiegazioni di tipo causale. Il modo più comune per fare accadere qualcosa è causarlo, e il modo più comune per spiegare qualcosa è specificarne le cause.

Poiché la forza esplicativa degli enunciati di ragione dipende dal modo in cui vengono descritti i fenomeni che riguardano la spiegazione, gli enunciati di ragione non sono estensionali. Il punto non è semplicemente il fatto che il connettivo “perché (because)” non è estensionale, ma il fatto che la sostituibilità non vale all'interno degli enunciati di ragione. Gli enunciati di ragione, in breve, sono intensionali-con-la-s non solo in relazione alla generalizzazione esistenziale, ma anche rispetto alla sostituibilità.

Consideriamo il seguente enunciato:

La California ha più terremoti di qualsiasi altro Stato perché la California è lo Stato con più faglie sismiche.

E consideriamolo ora assieme all'enunciato di identità:

Lo Stato con il maggior numero di faglie sismiche è lo Stato con il maggior numero di stelle del cinema

Questo non consente l'inferenza:

La California ha più terremoti di qualsiasi altro Stato perché la California è lo Stato con il maggior numero di stelle del cinema.

Il fallimento del criterio di sostituibilità all'interno di un tale enunciato di ragione è una conseguenza del fatto che la forza esplicativa della frase dipende da come vengono descritti i fenomeni in questione, dipende dalla forma aspettuale (*aspectual shape*) o dal modo di presentazione (*mode of presentation*). Se la specificazione dell'aspetto esplicativo – in questo caso l'aspetto causalmente effettivo – non viene conservata in seguito alla sostituzione delle espressioni di co-riferimento, allora la verità non viene conservata.

Anni fa vi furono accesi dibattiti riguardanti la questione se le ragioni sono cause. Ho sempre pensato che tali dibattiti fossero estremamente confusi perché non tenevano conto in maniera adeguata delle ovvie differenze grammaticali tra gli enunciati di ragione e gli enunciati causali. Le cause sono degli eventi, le ragioni non sono mai degli eventi. Si può ottenere una ragione enunciando una causa, ma questo non implica che la ragione e la causa siano la stessa cosa. Per rendere chiaro questo punto, utilizziamo un esempio.

(1) Perché l'autostrada sopraelevata di Oakland è crollata?

Questa domanda richiede una spiegazione e, perciò, una ragione. Solitamente si risponde specificando una causa, per esempio:

(2) Il terremoto Loma Prieta ha causato dei danni alle fondamenta.

L'enunciato (2) fornisce una ragione adeguata e perciò una spiegazione. Lo fa specificando una causa del crollo. Il terremoto, l'evento del danno alle fondamenta e l'evento del crollo dell'autostrada sono tre eventi collegati causalmente. Il terremoto ha causato il danno, il danno ha causato il crollo. (2) specifica questa sequenza e, perciò, è una spiegazione del terzo evento. L'enunciato di ragione, nell'enunciare i fatti, fornisce una spiegazione. La *causa* del crollo è un *evento*, il terremoto. La *ragione* del crollo è il *fatto* che c'è stato un terremoto che ha danneggiato le fondamenta. L'enunciato del fatto specifica la causa, ma la causa non è la stessa entità di una ragione.

Finora abbiamo fatto un piccolo progresso, anche se non è molto: le ragioni sono delle entità che hanno una struttura fattiva. La spiegazione è un atto linguistico che consiste nel fornire delle ragioni. Un enunciato di ragione sarà una spiegazione solo se la ragione stessa si trova in una, o più d'una, delle relazioni esplicative delle cose di cui è una ragione. Ma perfino questo piccolo progresso ha fornito un risultato interessante. Sebbene l'enunciato di ragione specifichi spesso una causa, non ne segue che in tali casi la causa sia identica alla ragione, perché le ragioni sono sempre delle entità fattive e le cause sono tipicamente degli eventi, non dei fatti.

II. ALCUNE CARATTERISTICHE SPECIALI DELLE SPIEGAZIONI DEI FENOMENI INTENZIONALI.

Quando introduciamo delle spiegazioni di fenomeni intenzionali, come le azioni, le credenze, le speranze così come le guerre, le politiche economiche, gli affari di cuore e i romanzi, noi introduciamo una nuova componente, la razionalità, e con la richiesta di spiegazioni razionali tipicamente arriva una richiesta di giustificazione. I fenomeni intenzionali sono soggetti ai vincoli della razionalità e la richiesta di spiegazione di un fenomeno intenzionale – una credenza, un desiderio, un'azione, ecc. – è caratteristicamente una richiesta di mostrare in che modo esso è razionale e in che modo esso è giustificato. Cioè, quando chiediamo una spiegazione con la domanda “perché l’hai fatto?” “perché credi questa cosa?”, “perché speri questa cosa?”, “perché vuoi questa cosa?” – così come “perché sei innamorato di lei?” “perché sei andato in guerra?” “perché hai abbassato il tasso d’interesse?” “perché hai scritto quel romanzo?” – noi introduciamo domande che non

appartengono solamente alla famiglia del “che cosa ha fatto sì che ciò accadesse?”, ma anche alla famiglia di “che giustificazione abbiamo per l’accadimento di questo?” e “in base a quale ragione hai agito?”. La razionalità degli stati intenzionali non è la stessa della giustificazione, poiché uno stato intenzionale può non avere una giustificazione senza per questo essere irrazionale. Potrei acquistare delle azioni al mercato dei cambi “in base a un’impressione”, dove la mia impressione non giustifica in alcun modo la mia scelta, ma la mia azione non risulta per questo necessariamente irrazionale. Sia la razionalità che la giustificazione sono nozioni normative, ma la razionalità è molto più generale della giustificazione. In generale gli stati intenzionali giustificati sono razionali, ma non tutti gli stati intenzionali sono giustificati.

Perché l’introduzione di ragioni esplicative per i fenomeni intenzionali introduce automaticamente le categorie normative della razionalità e della giustificazione? Perché è costitutivo dei fenomeni intenzionali essere soggetti a tali norme. *L’essere soggetti a dei criteri razionali di valutazione è costitutivo dei fenomeni intenzionali*, nello stesso modo in cui vincere e perdere sono costitutivi del gioco del calcio. Non si hanno prima le credenze, le speranze, i desideri e le intenzioni e, in seguito, al di fuori di queste, l’introduzione di forme razionali di valutazione; piuttosto, l’avere credenze ecc. significa già avere dei fenomeni che sono soggetti a queste norme. Inoltre, le diverse forme di intenzionalità hanno le loro forme proprie di normatività. Così, per esempio, si suppone che le credenze siano vere, e che per questa ragione esse siano soggette a speciali vincoli della razionalità e della giustificazione, che comportano, per esempio l’evidenza, altre ragioni per la verità, e la consistenza o coerenza. La razionalità richiede che una persona non possa consapevolmente avere delle credenze inconsistenti. La razionalità non ha tali richieste per i desideri: una persona può razionalmente volere che p e volere che non p .

Come di ogni altro fenomeno empirico del mondo reale, dei fenomeni intenzionali è possibile fornire delle spiegazioni causali lineari che non hanno nulla a che vedere con la razionalità o la giustificazione. Per esempio “Jones crede di essere Napoleone a causa di una commozione cerebrale”. Una spiegazione del genere è una spiegazione causale, ma non fornisce nessuna ragione che giustifichi la credenza di Jones o che dimostri che essa è razionale. Essa fornisce una ragione causale del perché egli ha quella credenza, ma non fornisce la *sua* ragione per quella credenza. La peculiarità dei fenomeni intenzionali è che essi sono, in virtù della loro stessa natura, anche soggetti ai vincoli della razionalità e, in quanto parte di quei vincoli, sono soggetti ad una richiesta di giustificazione.

Tutte le buone ragioni forniscono delle spiegazioni, e tutte le spiegazioni forniscono delle ragioni. Ma questo punto deve essere compreso in modo esatto. Si possono avere delle ragioni giustificatrici per credere qualcosa o per aver fatto qualcosa anche se l'enunciato di giustificazione non fornisce la ragione per cui si crede quel qualcosa o si è fatto quel qualcosa. *Le ragioni che giustificano la mia azione, e che così spiegano perché essa era l'azione giusta da compiere, possono non essere uguali alle ragioni che spiegano perché io, di fatto, ho compiuto quell'azione.* Così, se mi viene chiesto di giustificare il mio voto per Smith, potrei dire che ero giustificato a votare per lui perché era il candidato più intelligente. Ma fino a qui non ho risposto alla domanda sul perché ho votato per lui. Io potrei giustificare il mio voto dicendo che egli è il candidato più intelligente anche se la ragione *in base alla quale ho agito* è che egli era un mio vecchio compagno di bevute, e questo non ha nulla a che fare con l'intelligenza. In questo caso la giustificazione che io posso fornire della mia azione non è ancora una risposta alla domanda “perché l’hai fatto?”. Per prendere in considerazione un caso di maggior gravità, buona parte della discussione pubblica che riguardava il fatto se il Presidente Truman fosse giustificato a sganciare la bomba atomica, non riguarda le ragioni *in base alle quali egli agì*, ma riguarda la giustificazione dell'azione, se essa era una *cosa buona* sul piatto della bilancia. Tutti gli enunciati di ragione sono delle spiegazioni, ma il punto che sto affrontando ora è che la spiegazione del perché *qualcosa avrebbe dovuto essere fatto* o perché *è una buona cosa che esso sia stato fatto*, non è sempre la stessa cosa del *perché fu in effetti fatto*. In questo libro stiamo trattando, innanzitutto, delle spiegazioni che spiegano perché qualcosa è accaduto, spiegazioni che enunciano le ragioni in base alle quali l'agente ha agito o agirà. Ci interessano le giustificazioni solo in quanto spiegano anche perché l'agente ha agito o agirà. Distinguerò dunque tra le giustificazioni e quelle che chiamerò “spiegazioni giustificatorie”. La giustificazione non sempre spiega perché qualcosa di fatto è accaduto, ma una spiegazione del suo accadimento, giustificatoria o no, deve spiegare perché esso è accaduto. Una sottoclasse di spiegazioni genuine, è perciò costituita dalle spiegazioni giustificatorie.

Fino ad ora abbiamo rinvenuto quattro tipi di spiegazione degli stati intenzionali.

1. Le spiegazioni causali lineari. Esempio: Jones crede di essere Napoleone a causa di una commozione cerebrale.
2. Le spiegazioni di ragione del perché qualcosa è accaduto. Esempio: Jones ha votato per Smith perché Smith era un suo vecchio compagno di bevute.

3. Le spiegazioni giustificatorie. Esempio: Jones era giustificato a votare per Smith perché Smith era il candidato più intelligente e questa è la ragione per la quale ha votato per lui.
4. Le giustificazioni che non sono spiegazioni del perché l'azione è accaduta. Esempio: Jones era giustificato a votare per Smith perché Smith era il candidato più intelligente, anche se questa non è la ragione per la quale, di fatto, ha votato per lui.

Tenendo presente tutto questo, vorrei ora esprimere un punto cruciale: *l'introduzione di vincoli normativi nelle spiegazioni di ragione del perché un qualche fenomeno intenzionale è accaduto non rimuove i vincoli causali*. A causa del divario, le cause delle azioni e di molti altri fenomeni intenzionali, normalmente, non forniscono condizioni sufficienti, così, con una formulazione più precisa dovremmo dire: *per quanto riguarda i fenomeni intenzionali, i vincoli normativi imposti alla spiegazione del perché un'azione è accaduta, del perché un agente ha accettato una credenza, perché un agente ha formato un desiderio, perché un agente si è innamorato ecc. non rimuovono il vincolo causale secondo il quale una spiegazione del perché l'agente ha fatto ciò che ha fatto deve esprimere le ragioni che erano effettive per l'agente stesso*. Si possono avere spiegazioni causali di fenomeni intenzionali non razionali, ma non si possono avere spiegazioni razionali del perché un certo fenomeno intenzionale è accaduto che non contengano la nozione di effettività causale. Nel caso delle azioni, l'agente rende effettiva una ragione agendo in base ad essa. Nel caso della credenza l'agente accetta la credenza a causa anche di una ragione che ha accettato. Nel caso dei desideri motivati l'agente forma il desiderio sulla base di una ragione. Così, per esempio, alla domanda "perché hai votato per il candidato democratico?", qualcuno potrebbe rispondere "E' semplicemente un'ossessione irrazionale che ho. Non posso farci nulla, sono sempre costretto a votare per i Democratici". Una spiegazione del genere fornisce una spiegazione causale, ma non razionale, e ancor meno giustificatoria. Ma se qualcuno dicesse "ho votato per il candidato democratico perché i Democratici sosterranno maggiormente i sindacati dei lavoratori, ed io sono impegnato a sostenere i sindacati dei lavoratori" questa spiegazione, per fornire una spiegazione razionale dell'azione, deve essere anche una spiegazione causale. L'agente *agisce* in base alla credenza e al suo impegno. Si possono dare giustificazioni di fenomeni intenzionali che non sono causali ma finché la giustificazione non enuncia una ragione che sia causalmente effettiva, essa non fornisce una spiegazione del perché sono accaduti i fenomeni

intenzionali. Così come per le azioni, questo è vero anche per credenze, i desideri e le emozioni.

Per riassumere: fino ad ora ho sostenuto tre tesi sostanziali. Per prima cosa, che tutte le ragioni sono entità fattive che stanno in una o più relazioni esplicative con le cose delle quali esse sono ragioni. Secondo, che i fenomeni intenzionali sono, in aggiunta, soggetti a vincoli normativi. Terzo, se stiamo spiegando perché qualcuno ha fatto qualcosa, o ha un qualche fenomeno intenzionale, questi vincoli normativi non rimuovono il vincolo causale. Ragioni e razionalità, per fornire delle spiegazioni, devono funzionare causalmente (tenendo conto del divario, naturalmente). La peculiarità dei fenomeni intenzionali è che essi ammettono sia spiegazioni causali non normative che spiegazioni causali normative. *Ma le spiegazioni normative, per spiegare l'accadimento del fenomeno intenzionale, devono essere anche causali.* I fenomeni non intenzionali, come i terremoti, ammettono soltanto spiegazioni non normative. Per questa ragione, le giustificazioni di un fenomeno intenzionale non sono sempre delle spiegazioni del perché esso è accaduto. Così, per ripetere, abbiamo almeno quattro tipi di casi. Per prima cosa, le spiegazioni causali non intenzionali: ad esempio, egli crede di essere Napoleone a causa di una commozione cerebrale. Secondo, le spiegazioni razionali del perché un fenomeno è accaduto, che non intendono fornire giustificazione. Terzo, le giustificazioni del perché un fenomeno è accaduto che spiegano anche perché è accaduto. E quarto, le semplici giustificazioni che non spiegano perché un fenomeno è accaduto.

III. RAGIONI PER L'AZIONE E RAGIONI TOTALI

Fino ad ora, questo capitolo è stato rivolto ad una chiarificazione preliminare. Dobbiamo ora passare alla parte costruttiva. Il cuore dell'argomento di questo capitolo è in questa parte, e poiché aspiro ad una chiarezza totale, esporrò l'argomento in una serie di passi numerati. Inizierò con alcuni dei punti che abbiamo analizzato nelle due parti precedenti.

1. Le ragioni sono sia proposizionali che relazionali. Per essere una ragione un'entità deve avere una struttura proposizionale e deve essere correlata a qualcos'altro che abbia anch'esso una struttura proposizionale e per il quale essa sia una ragione. Così, tutte le

ragioni sono tali solo in relazione alle cose di cui sono ragioni. Questo punto banale e grammaticale porta con sé la conseguenza che, per quanto riguarda l'intenzionalità, una ragione è sempre una ragione per uno stato intenzionale. E' una ragione per credere ad una proposizione o una ragione per avere un desiderio, o una ragione per formare un'intenzione precedente, o una ragione per un'intenzione-in-azione, cioè, una ragione per compiere effettivamente l'azione. Nel caso speciale delle ragioni per l'azione, una ragione è anche una ragione perché una persona compia un'azione e, se la ragione deve funzionare nella deliberazione, essa deve essere conosciuta dall'agente.

2. Le ragioni sono delle entità fattive. Le ragioni per la mia azione possono essere fatti del mondo, come il fatto che sta piovendo, o possono essere stati intenzionali con una struttura fattiva, come le credenze e i desideri, o possono essere entità fattive del mondo come i doveri, gli obblighi e gli impegni, dotate tutte di una direzione di adattamento verso l'alto.

3. E' necessario distinguere tra ragioni interne e ragioni esterne. Una ragione esterna, nel senso in cui io la intendo, è un'entità fattiva del mondo che può essere una ragione per un agente, anche se egli non conosce quell'entità, o la conosce ma rifiuta di riconoscerla come una ragione. Così, il fatto che sta piovendo, o il fatto che una persona ha un obbligo, è una ragione esterna. Affinché una tale ragione esterna funzioni nella deliberazione effettiva, essa deve essere rappresentata da un qualche stato intenzionale interno dell'agente. L'agente *crede* che stia piovendo, o l'agente *riconosce* il suo obbligo. Così in una situazione razionale ideale, c'è un adattamento tra le ragioni interne e le ragioni esterne, perché fintanto che vi sono delle ragioni esterne che giocano un ruolo nella deliberazione, esse saranno rappresentate come ragioni interne nella mente dell'agente. La deliberazione dell'agente può operare soltanto sulla base di ragioni interne, ma le ragioni interne spesso sono ragioni valide solo perché rappresentano delle ragioni esterne. Così, per esempio, se decido di prendere un ombrello perché credo che stia piovendo, la mia credenza è una ragione interna ma è una ragione valida solo se corrisponde ad una ragione esterna, solo se, di fatto, sta piovendo.

4. Una ragione per un'azione è una ragione solo se essa è, o è parte, di una ragione totale. Ho sostenuto che le ragioni per l'azione sono relative in almeno tre modi, ma c'è un quarto modo che richiede anch'esso di essere evidenziato: un enunciato è un enunciato di una *ragione per un'azione* solo fino a che quell'enunciato è sistematicamente correlato a certi altri enunciati. Questo può essere osservato considerando degli esempi. La mia ragione

per prendere un ombrello è che io credo che stia per piovere. Ma la mia ragione è tale solamente perché fa parte di una ragione totale che comprende cose come il mio desiderio di rimanere asciutto e la mia credenza che se ho un ombrello rimarrò asciutto.

Una ragione totale è un insieme di entità fattive. Queste possono essere delle credenze, dei desideri o dei fatti del mondo come il fatto che sta piovendo o il fatto che sono obbligato ad andare a Kansas City. Così, in risposta alla domanda “perché prendi un ombrello?” posso dire cose come “sta per piovere”, “credo che stia per piovere” o “non voglio bagnarmi”.

5. Una ragione totale potrebbe, per principio, essere interamente esterna. Per esempio, qualcuno potrebbe avere una ragione per mangiare degli agrumi senza avere nessuno stato intenzionale rilevante. Così, supponiamo sia un fatto che gli agrumi contengano vitamina C; la vitamina C previene lo scorbuto; lo scorbuto è una malattia terribile. Tutti questi potrebbero essere gli elementi di una ragione totale per mangiare degli agrumi persino per chi non sa nulla di agrumi o è indifferente riguardo alla malattia.

In che senso si può dire che una ragione totale interamente esterna sia una ragione per un agente, se non può in alcun modo motivare l’agente? La risposta è che la forza motivazionale di una ragione esterna è definita in modo controfattuale: se l’agente aveva una conoscenza appropriata, cioè, se era a conoscenza delle esigenze della sua salute, e sapeva come soddisfarle, allora, se razionale, egli le avrebbe riconosciute come ragioni per un’azione. Così, sebbene vi sia un accordo ideale tra le ragioni interne e quelle esterne, noi abbiamo ancora bisogno di una distinzione tra le due. Un agente perfettamente razionale potrebbe agire razionalmente sulla base di una credenza razionalmente giustificata che si rivelasse falsa, e un fatto del mondo potrebbe essere una ragione irresistibile perché un agente agisca, persino in casi in cui l’agente non avesse una conoscenza del fatto in questione, o ne fosse a conoscenza ma si rifiutasse di riconoscerlo come una ragione.

6. Per poter funzionare nella deliberazione razionale e nei processi razionali che portano all’azione, ogni elemento di una ragione totale esterna deve essere in accordo con un elemento interno. Cioè, i fatti che costituiscono una ragione esterna devono essere creduti, conosciuti, riconosciuti o in altri modi resi oggetto di consapevolezza da parte dell’agente in questione. Così una necessità della salute, o un obbligo, o il fatto che stia piovendo possono funzionare nella deliberazione che motiva un’azione solamente se l’agente in questione crede o riconosce in qualche modo il fatto in questione. Il fatto che stia per piovere, può essere una ragione perché io prenda un ombrello, nonostante io non conosca quel fatto. Ma

il fatto che stia per piovere può giocare un ruolo nelle mie deliberazioni solamente se sono consapevole di quel fatto. Inoltre, la credenza che sta per piovere giocherà lo stesso ruolo nella mia deliberazione sia se la credenza è vera sia se non è vera. Questo può far pensare che ciò che davvero importa non sia il fatto stesso ma la credenza. Ma questo è errato. La credenza è ciò che risponde ai fatti. Infatti, in alcuni casi, la razionalità può richiedere una credenza piuttosto che un'altra. Così, se guardo fuori dalla finestra e vedo che sta piovendo, sarebbe irrazionale, a parità di altre cose, rifiutare di credere che sta piovendo.

Potrebbe sembrare che esista la minaccia di un regresso all'infinito: la razionalità richiede la credenza, ma l'acquisizione della credenza in sé richiede la razionalità. Perché questo non porterebbe ad un regresso all'infinito?

7. Per mostrare perché tali casi non portano ad un regresso all'infinito, ho bisogno di introdurre la nozione di *razionalità riconoscitiva*. La razionalità può richiedere che un agente, a certe condizioni epistemiche, riconosca semplicemente un fatto nel mondo come il fatto che egli ha accettato un obbligo, o ha una particolare esigenza, o che egli si trova in un certo tipo di pericolo ecc. anche se non c'è alcun processo razionale, nessuna attività di deliberazione, che porti al risultato razionale. L'acquisizione di uno stato intenzionale razionale non richiede sempre un processo razionale di deliberazione, o, addirittura, non richiede nessun tipo di processo.

Possiamo vedere che queste acquisizioni sono razionali mettendole a confronto con le loro negazioni irrazionali. Infatti, una forma comune di irrazionalità è la "negazione", quando un agente nega insistentemente qualcosa di fronte a una schiacciante evidenza. Per esempio, tempo fa avevo un amico che è diventato un alcolista. Per molto tempo egli ha rifiutato insistentemente di riconoscere che era un alcolista. Egli pensava semplicemente che gli piacesse bere un po' di più delle altre persone. Altri esempi sono i casi in cui le persone semplicemente si rifiutano di riconoscere gli obblighi che hanno sottoscritto, o si rifiutano di credere di essere stati traditi o di essere in pericolo. L'elemento che emerge da tali esempi è che le attitudini irrazionali prendono avvio da un semplice riconoscimento razionale dei fatti. Ma il riconoscimento razionale dei fatti non richiede necessariamente la deliberazione. Io posso semplicemente guardare e vedere che un camion mi sta investendo, o guardare fuori dalla finestra e vedere che sta piovendo. In entrambi i casi, io riconosco che questi fatti mi forniscono delle ragioni per l'azione. In questi casi la razionalità richiede che io creda che stia piovendo o che il camion mi stia investendo, ma non che io mi impegni in un processo di deliberazione razionale per raggiungere queste conclusioni razionali. Molte

ragioni interne sono basate sul riconoscimento di una ragione esterna. Il riconoscimento razionale di una ragione esterna in molti casi non richiede nessuna ulteriore deliberazione. La razionalità riconoscitiva non richiede necessariamente che si passi attraverso numerosi passaggi.

8. L'insieme degli elementi fattivi che costituiscono una ragione totale deve contenere almeno un elemento che abbia la direzione di adattamento mondo-a-mente. Chiamiamo questi elementi che hanno una direzione di adattamento mondo-a-mente e che sono, almeno potenzialmente, capaci di funzionare nelle ragioni totali, *motivatori*: *ogni ragione totale deve contenere almeno un motivatore*. Perché? Perché la razionalità, nella deliberazione che riguarda le azioni, è una questione di trovare dei modi per soddisfare i motivatori. L'argomento più semplice per la tesi secondo la quale una ragione totale deve contenere almeno un motivatore è che una ragione totale deve essere capace di motivare razionalmente un agente. Una ragione totale deve fornire uno sfondo razionale per un'intenzione precedente per compiere l'azione o per la realizzazione intenzionale dell'azione. Per poter fare questo, deve esserci una qualche entità nella ragione totale che abbia la direzione di adattamento mondo-a-mente, e che faccia da sfondo per la direzione di adattamento mondo-a-mente dell'intenzione precedente e dell'intenzione-in-azione.

Nel caso in cui il motivatore sia un fatto del mondo epistemicamente oggettivo, come il fatto che l'agente ha certe necessità o certi obblighi, il motivatore esterno può funzionare nella deliberazione solo se è riconosciuto tale dall'agente. E, per ripetere il punto che ho sottolineato nel paragrafo precedente, la razionalità riconoscitiva può richiedere che l'agente riconosca il motivatore come un motivatore. L'uomo che rifiuta di riconoscere che c'è un camion che sta investendolo, mettendosi in grave pericolo fisico è, in quel caso, semplicemente irrazionale, anche se egli non ha effettuato alcun processo di deliberazione. Ma il punto cruciale di questa discussione è che affinché i motivatori esterni funzionino nella deliberazione, essi devono essere riconosciuti come tali dall'agente.

I motivatori possono essere sia esterni che interni. I desideri, per esempio, sono dei motivatori interni, i bisogni e gli obblighi sono dei motivatori esterni. Ma, ripeto, i motivatori esterni possono funzionare nella deliberazione solamente quando vengono rappresentati come dei motivatori interni. Una ragione per l'azione totale interna deve contenere almeno un motivatore riconosciuto come tale.

9. Il requisito per cui il ragionamento deve avere un motivatore è vero sia per la ragione teorica che per la ragione pratica. Supponete dunque che io creda a proposizioni della forma

che p e che se p allora q . Che cosa ha a che fare tutto questo con il mio accettare, riconoscere o credere che q ? Se le credenze sono solo degli oggetti neutri, degli insiemi di relazioni causali, secondo una teoria molto in voga (ma sbagliata), per quale ragione io, questo sé, dovrei prendere in considerazione q ? La risposta è che una credenza è un impegno nei confronti della verità. E quando ho una credenza io sono impegnato nei confronti di tutte le sue conseguenze logiche. E un impegno è un motivatore esterno indipendente dal desiderio che ha direzione di adattamento mondo-a-mente. Questa è la ragione reale del perché non esistono, in questo senso, delle distinzioni di principio tra la ragione pratica e quella teorica. La ragione teorica è quella branca della ragione pratica che riguarda le ragioni per accettare, riconoscere, credere e asserire delle proposizioni.

10. L'elenco dei motivatori sembra, a prima vista, così eterogeneo da risultare scoraggiante. Include motivatori interni come il desiderio, la speranza, la paura, la vergogna, l'orgoglio, il disgusto, l'onore, l'ambizione, l'amore e l'odio, per non menzionare la fame, la sete e la sensualità. Include motivatori esterni come i bisogni, gli obblighi, gli impegni, i doveri, le responsabilità e le richieste. Notate che entrambi questi insiemi di motivatori sono fattivi, nel senso che ho spiegato prima.

11. I motivatori esterni sono entità fattive del mondo. Sotto le descrizioni che li identificano come motivatori esterni, descrizioni date in termini di "bisogno", "obbligo", "impegno", "richiesta", "dovere" ecc. essi sono sempre relativi-all'osservatore. È solamente in relazione all'intenzionalità umana che qualche stato di cose nel mondo, per esempio, può essere identificato come una necessità per la salute. La relatività nei confronti dell'osservatore implica la soggettività ontologica, ma essa non implica necessariamente la soggettività epistemica. Questo significa che l'ontologia dei fenomeni relativi-all'osservatore contiene sempre un qualche riferimento all'intenzionalità degli osservatori in questione. Quindi l'ontologia è soggettiva. Ma è quasi possibile per gli enunciati che riguardano entità ontologicamente soggettive avere un'oggettività epistemica. Può essere un fatto oggettivo che io abbia una certa necessità per la salute, sebbene la sua identificazione come "necessità" è relativa-all'osservatore.

Questo è un punto importante, quindi affrontiamolo con un esempio. Supponiamo che nel mio corpo io abbia un certo livello di vitamina C. Questo è semplicemente un fatto brutto che mi riguarda, ed è indipendente-dall'osservatore. Ma supponiamo che il livello sia tale da risultare insufficiente per prevenire le malattie, quindi

(a) ho bisogno di più vitamina C.

Ora, quale fatto corrisponde all'affermazione che io *ho bisogno* di più vitamina C? Quali fatti sono costitutivi di quel fatto? I fatti bruti del mondo sono cose come il fatto che ho un certo livello di vitamina C nel corpo, che il mio corpo ha certi processi causali, e che il livello di vitamina C è insufficiente a mantenere questi processi. Insieme questi fatti costituiscono il bisogno, ma sotto la descrizione di “bisogno”, o “necessità”, questi fatti hanno una direzione di adattamento verso l’alto. Questo è dimostrato dal fatto che un bisogno può essere colmato o soddisfatto, ma non può essere vero o falso. Un bisogno è colmato o soddisfatto se e solo se il mondo arriva ad adattarsi al contenuto proposizionale del bisogno. Il fatto bruto del mondo, il fatto che io abbia un certo livello di vitamina C, non ha una direzione di adattamento. Ma il fatto è sufficiente a costituire un motivatore relativo-all’osservatore: ho bisogno di più vitamina C. E sotto la descrizione di “bisogno” il fatto è un motivatore capace di funzionare come ragione per l’azione.

L’enunciato (a) enuncia un fatto che è una ragione per l’azione. Questa ragione è un motivatore esterno, il mio bisogno. I bisogni sono relativi all’osservatore. E' soltanto in relazione alla mia salute e alla mia sopravvivenza che io ho un bisogno come questo. Anche se il bisogno è relativo-all’osservatore ed è dunque ontologicamente soggettivo, è un fatto epistemicamente oggettivo che riguarda me stesso il fatto che io abbia un tale bisogno: cioè non è semplicemente una questione di *opinione* il fatto che io abbia questo bisogno, ma è un chiaro fatto medico di tipo oggettivo.

12. I motivatori-indipendenti dal desiderio, sotto la loro descrizione come motivatori, hanno sempre la direzione di adattamento verso l’alto, mondo-a-motivatore. Per questa ragione il loro riconoscimento sotto queste descrizioni, ovvero il loro riconoscimento come motivatori, è già un loro riconoscimento come ragioni per l’azione. L’agente non deve, per prima cosa, riconoscere un obbligo e poi comprendere che ha una ragione per l’azione, perché riconoscere qualcosa come un obbligo è già riconoscerla come un motivatore nel senso che abbiamo spiegato.

13. La razionalità nel prendere decisioni, include almeno i tre seguenti elementi. Per prima cosa, il riconoscimento dei vari motivatori, sia esterni che interni, e la valutazione dei loro pesi relativi. Supponiamo che io abbia promesso di venire alla vostra festa il prossimo mercoledì sera. Ho chiaramente un obbligo di venire alla vostra festa, e tale obbligo è una ragione indipendente dal desiderio, non ha nulla a che fare con il mio desiderio di venire alla

vostra festa. Ma supponiamo anche che sia fortemente contrario ai miei interessi venire alla vostra festa, poiché se verrò perderò un affare di lavoro che mi costerà un'intera fortuna. Quell'interesse è contrario al motivatore esterno con la cui forza bisogna fare i conti. Spesso i filosofi morali, Kant per esempio, hanno sostenuto che in simili casi di interesse personale contrario al dovere, il dovere dovrebbe sempre trionfare. Ma questo mi sembra semplicemente ridicolo. Ci sono molti casi in cui io ho un obbligo minore, come il mio obbligo di venire alla vostra festa, ed ho anche interessi molto profondi che sono in conflitto con quell'obbligo. Non c'è una ragione perché il motivatore indipendente dal desiderio debba sempre trionfare.

Secondo, deve esserci un riconoscimento corretto e una valutazione dei fatti non motivazionali che riguardano il caso. Così, per esempio, io devo essere capace di sapere in che modo sarò in grado di assolvere i miei diversi obblighi. È almeno fisicamente possibile per me soddisfare tutti gli obblighi verso i quali mi sono impegnato? Per semplificare, possiamo dividere questi fatti non motivazionali in due tipi. Quelli che hanno a che fare con le relazioni del tipo “tramite” e “per mezzo di” o “attraverso il compimento (di un'altra azione)”, nel senso che ho spiegato nel capitolo 2. Questi sono fatti che riguardano il come soddisfare i motivatori e il che cosa costituisce il soddisfare i motivatori. Li chiameremo rispettivamente *effettuatori* e *costitutori*. E, ancora una volta, dobbiamo distinguere tra effettuatori e costitutori interni ed esterni. Un semplice esempio renderà chiare queste distinzioni. Supponiamo che io vi debba del denaro (motivatore esterno). Supponiamo che io lo sappia (motivatore interno). Supponiamo che io possa saldare questo debito venendo in automobile fino a casa vostra e dandovi il denaro (effettuatore e costitutore esterni). Sapendo tutte queste cose io posso decidere di venire in automobile fino a casa vostra e darvi il denaro (ragione pratica).

Gli effettuatori e i costitutori interni sono sempre delle credenze. Sono credenze che riguardano come compiere le cose causalmente (effettuatori) o come fare una certa cosa costituisca il fare qualcos'altro (costitutori). In quanto credenze, gli effettuatori e i costitutori interni sono responsabili di come stanno le cose nel mondo reale. Essi hanno una direzione di adattamento verso il basso. Così essi sono ragioni valide per l'azione solo quando corrispondono ai fatti reali del mondo. Il fatto che io possa sparare con la pistola premendo il grilletto è un effettuatore esterno. Quindi, se io ho una ragione per sparare con la pistola, allora ho una ragione per premere il grilletto. L'effettuatore esterno sarà effettivo

all'interno del mio ragionare solo se esiste un esecutore interno corrispondente, la mia credenza che premendo il grilletto io posso sparare con la pistola.

È questa combinazione di caratteristiche, ovvero l'esistenza dei motivatori e il riconoscimento dei fatti su cui si fonda il caso, che dà alle persone l'illusione che, in qualche modo, il ragionare significhi un ragionare in termini di mezzi-fini o di credenze-desideri. I motivatori forniscono i fini (desiderati) e i fatti non motivazionali forniscono i mezzi (creduti). Ma questo modo di guardare le cose offusca la distinzione tra motivatori interni ed esterni, e, in modo analogo, anche la distinzione tra ragioni per l'azione dipendenti dal desiderio e ragioni indipendenti dal desiderio. Il grande abisso che esiste tra gli esseri umani e gli scimpanzé, per quanto riguarda la ragione pratica, è che noi abbiamo la capacità di creare, di riconoscere e di agire sulla base di ragioni per l'azione indipendenti dal desiderio. Nella storia della filosofia occidentale il grande rompicapo della razionalità è sempre stato: com'è possibile che un agente possa essere razionalmente motivato da ragioni indipendenti dal desiderio? Dato che ogni azione è in un certo senso l'espressione del desiderio di compiere quell'azione, da dove proviene allora il desiderio, se la ragione in base alla quale l'agente agisce non è né essa stessa un desiderio, né è essa stessa basata su altri desideri? Come possono le ragioni indipendenti dai desideri fornire razionalmente la base per un desiderio? La risposta tradizionale a queste domande, fornita dal Modello Classico, è che l'agente deve avere un qualche desiderio dominante, o di ordine superiore, per agire in base a queste ragioni indipendenti dal desiderio. Così l'agente deve avere qualche desiderio generale di dire la verità o di mantenere le sue promesse o di assolvere i suoi obblighi. Ma questo deve essere il modo sbagliato di considerare questi argomenti, perché implica che nei casi in cui l'agente non ha questi desideri di ordine superiore, egli non ha alcuna ragione per dire la verità, assolvere i suoi obblighi o mantenere le sue promesse. Ciò che dobbiamo dimostrare è come il semplice fatto che un agente riconosca qualcosa come una asserzione, una promessa, o un'altra forma d'obbligo sia già una base per una motivazione. Come è possibile questo? La risposta breve è che tutte queste cose hanno una direzione di adattamento verso l'alto, e riconoscere che certi tipi di entità fattive hanno una direzione di adattamento verso l'alto e che in esse l'agente è il soggetto del contenuto proposizionale, significa già riconoscere una ragione per agire sulla base di quel contenuto proposizionale. Discuterò ulteriormente questo punto nel capitolo 6.

Il terzo elemento presente nel prendere decisioni razionali, una volta che si sia assemblata la ragione totale, è valutare l'insieme dei motivatori e dei fatti non motivazionali

in modo da arrivare ad una decisione. La teoria della decisione mi sembra fornisca al riguardo una spiegazione notevolmente superficiale perché essa assume che io abbia uno schema ben ordinato di preferenze preesistente e che si tratti semplicemente della capacità di elaborare delle stime di probabilità su come giungere al gradino più alto della mia scala di preferenze. Ma la vera difficoltà sta nel costruire lo schema di preferenze. Gran parte della difficoltà nella deliberazione razionale è decidere che cosa si vuole davvero, e che cosa davvero si vuole fare. Non si può assumere che l'insieme dei voleri sia ben ordinato precedentemente alla deliberazione. Inoltre dobbiamo considerare che non tutti i motivatori sono sullo stesso livello.⁴

Secondo il Modello Classico noi assumiamo che l'insieme dei fini sia dato precedentemente alla deliberazione. Questi fini sono tutti, in generale, cose che l'agente desidera. La deliberazione riguarda allora la selezione dei mezzi per questi fini, riguarda i modi per soddisfare i desideri. Nella gran parte delle spiegazioni, si assume che l'insieme dei desideri sia consistente. Secondo la spiegazione alternativa che sto qui proponendo, tutto questo è irrimediabilmente errato. La parte più importante della ragione pratica è data dallo scoprire quali sono in primo luogo i fini. Alcuni di questi sono dei desideri, ma altri sono ragioni per l'azione indipendenti dal desiderio, razionalmente irresistibili. *Per questo la ragione è la base del desiderio; non è il desiderio ad essere la base della ragione.* Ovvero, quando vi rendete conto che avete una ragione per fare qualcosa voi non volete fare altrimenti, dovete farlo *a fortiori*, dovete volerlo fare. E, qualche volta, ma non sempre, quel riconoscimento vi porterà a volerlo fare.

Inoltre, persino dopo che avete individuato i vostri motivatori, cioè le vostre ragioni per l'azione, sia dipendenti che indipendenti dai desideri, l'insieme sarà ancora consistente. Non potete fare tutte le cose che vorreste fare, o tutte le cose che dovete fare. Dovrete dunque avere una qualche capacità di valutare la forza relativa dei motivatori. Ma anche quando riuscirete a risolvere quel problema, per la vostra soddisfazione razionale, ancora non avrete una chiara distinzione fini-mezzi, poiché alcuni dei mezzi comprendono i loro fini e alcuni mezzi interferiscono con altri fini. Per fare l'esempio più semplice, se uno dei

⁴ C'è un aneddoto che riguarda un famoso teorico della decisione. Gli fu offerto un lavoro allettante presso un'altra università, un lavoro che lo tentava, sebbene fosse profondamente impegnato con l'università presso la quale lavorava. Egli discusse con un amico se accettare o no. Il suo amico gli disse che, dal momento che era un famoso teorico della decisione, egli doveva essere capace di applicare la sua teoria della decisione per prendere questa

vostri scopi, o fini, è quello di risparmiare del denaro, troverete che i mezzi per raggiungere molti altri vostri fini implicano proprio lo spendere del denaro.

Cercherò di rendere tutto questo più chiaro nelle pagine seguenti, ma in questo momento mi soffermerò su alcuni esempi.

IV. PRENDERE DECISIONI NEL MONDO REALE

In un caso ordinario, come il mio tentativo di organizzare il tempo per scrivere questo libro, io ho una serie di motivatori in conflitto che riguardano la situazione. Ho l'obbligo di finire questo libro. Ma ho altri obblighi che riguardano altri scritti che devono essere completati prima di questo. Io considero questo lavoro il più importante ed ho promesso di terminare il manoscritto entro una data che è troppo vicina. Il mio obbligo di scrivere questo libro è in conflitto con l'obbligo di produrre altri due articoli che devo scrivere entro la fine di questo mese. D'altro canto ho un'idea piuttosto vaga di che cosa fare di questo manoscritto, e gli altri obblighi sembrano più semplici da portare a termine. Mi aspetto di essere pagato di più per questo libro che per gli articoli. Ho anche l'insegnamento e degli impegni famigliari che devono assolutamente essere assolti. Per esempio devo tenere le lezioni nei miei corsi universitari e devo essere a casa per la cena. Fare filosofia è gratificante e lo sono pure le molte altre cose da fare, ma non posso farle tutte.

Questa è la ragione pratica nella vita reale. Notate: non posso fare una chiara distinzione tra doveri e desideri, né tra scopi e mezzi. Generalmente non avrei questi doveri se non avessi voluto averli e se non avessi voluto fare le cose che mi obbligano a fare. I miei desideri hanno prodotto questi doveri. Scrivere questo libro è uno scopo o un mezzo? La risposta è che è entrambi, in differenti modi. Ma qual è la massima della mia azione? E non dovrei cercare di vedere se essa può valere come una legge universale? Ancora, io posso formare una quantità di massime differenti, alcune possono diventare universali, alcune no, e ciò non sembra molto importante. L'idea che per essere un agente razionale dovrei avere innanzitutto uno schema ben ordinato di preferenze e poi fare delle valutazioni probabilistiche in relazione a quale corso di azione potrebbe massimizzare la mia aspettativa utilitaristica, sembra un'idea priva di qualsiasi plausibilità.

decisione. Ciò che l'amico non sapeva è che la teoria della decisione si applica, in gran parte, solamente dopo che le parti principali della decisione sono già state realizzate.

Ma in tutto questo apparente caos intenzionale, c'è effettivamente un ordine, e lo scopo della ragione pratica è rendere chiaro ed estendere questo ordine.

Ecco il primo serio rompicapo: come possono i fatti del mondo, come il fatto che io ho un certo livello di vitamine o il fatto che ho pronunciato certe parole, costituire un motivatore razionalmente irresistibile? Bene, alcuni di questi fatti, in base a certe descrizioni, sono già dei motivatori. Così quella frase era una promessa e quindi il sottoscrivere un obbligo. Quel livello di vitamine è una mancanza, e perciò un bisogno. La razionalità riconoscitiva può richiedere che io riconosca le mie mancanze e i miei bisogni sotto queste descrizioni e, perciò, li riconosca come motivatori. Ma in che modo? Non ho bisogno di qualche altro desiderio antecedente per mantenere i miei obblighi o per soddisfare i miei bisogni sanitari? Ho detto prima che i principi della razionalità riconoscitiva possono richiedere che certi fatti esterni siano riconosciuti come motivatori esterni e che siano così rappresentati come motivatori interni. Ma bisogna dire qualcosa di più circa i principi della razionalità riconoscitiva. Ho detto che non ci sarebbe nessun regresso all'infinito. Ma perché no? Non avrei bisogno di un motivatore per il motivatore? E questo non porterebbe a un altro tipo di regresso all'infinito?

La semplice verità è che posso impegnarmi in un ragionamento solo in relazione a ciò che è interno alla mia mente e questo non è incoerente con la tesi che il riconoscimento dei fatti oggettivi del mondo possono essere richiesti razionalmente ed insieme fornire una base razionali per i motivatori interni.

V. COSTRUIRE UNA RAGIONE TOTALE: UN TEST PER IL MODELLO CLASSICO

Assumendo dunque che una ragione totale debba contenere questi tre tipi di elementi, in che modo, esattamente, noi costruiamo, valutiamo e agiamo sulla base di una ragione totale? Voglio considerare un caso tratto dalla vita reale, perché illustra la differenza tra la prospettiva che sto proponendo e il Modello Classico. Credo che l'esempio che sto per esporre sia un esempio di irrazionalità, tuttavia il Modello Classico non può descrivere tale irrazionalità.

Quando insegnavo in Danimarca avevo una studentessa che fumava moltissime sigarette. Le feci notare che fumare era molto dannoso per la sua salute. E lei era d'accordo. Ebbene, dissi "perché allora continui a fumare?" Lei disse che non si curava della sua salute,

che era felice di morire molto più giovane di quando sarebbe morta se non avesse fumato, ma in quel momento voleva fumare. In quel momento voleva assolutamente fare qualcosa che sapeva avrebbe avuto la conseguenza di portarla alla morte all'età di sessant'anni. Le feci notare che quando avesse avuto sessant'anni non avrebbe voluto morire all'età di sessant'anni e avrebbe rimpianto di aver fumato in quel momento. Lei era d'accordo che sì, a sessant'anni non avrebbe voluto morire all'età di sessant'anni, e avrebbe rimpianto di aver fumato a vent'anni, ma proprio in quel momento, all'età di vent'anni, quando doveva prendere la decisione, era assolutamente felice di morire all'età di sessant'anni, ed era proprio allora il momento in cui doveva prendere la decisione se fumare o no.

L'aspetto interessante del caso è che lei era d'accordo con tutti i fatti che le facevo notare. Era d'accordo che fumare corrispondeva, molto probabilmente, ad uccidere se stessa all'età di sessant'anni, che avvicinandosi a quell'età avrebbe rimpianto di aver fumato e che allora non le sarebbe andato di morire per colpa del fumo, ma ciononostante proprio in quel momento, dovendo prendere una decisione se fumare o no proprio in quel momento, la cosa razionale per lei era fumare, perché voleva fumare proprio in quel momento. Cioè, lei non ammetteva nessuna forma di irrazionalità. Al contrario, insisteva che il suo comportamento era totalmente razionale, che per lei la cosa razionale da fare proprio in quel momento era fumare.

Secondo il Modello Classico la sua azione era in effetti totalmente razionale. Le sue credenze e i suoi desideri erano tali che fumando lei otteneva la massima soddisfazione dei suoi desideri, date le sue credenze. È vero che lei avrebbe potuto avere dei desideri ulteriori che non sarebbero stati soddisfatti, ma quei desideri ulteriori non potevano giocare alcun ruolo nel prendere la decisione razionale nella quale era impegnata proprio in quel momento, poiché quei desideri ulteriori non esistevano in quel momento. Inoltre, non aveva desideri di secondo ordine presenti riguardo a quei desideri futuri; essi le erano totalmente indifferenti. Lei non pensava “desidererò quella tal cosa in futuro, quindi desidero desiderarla ora”. I desideri futuri non giocavano alcun ruolo per lei.

Secondo la versione di Bernard Williams del Modello Classico, noi dovremmo dire che il suo era un caso di perfetta razionalità, perché ella agiva soltanto sulla base di ragioni interne, e le ragioni interne non includevano alcuna preoccupazione riguardo ai suoi futuri quarant'anni. Qualsiasi cosa avessi detto per spingerla a smettere di fumare si sarebbe richiamato ad una ragione esterna, qualcosa di esterno al suo insieme motivazionale attuale, e per quella ragione, secondo il modello di Williams, non avrebbe potuto avere alcun effetto

sulla sua razionalità. C'era, secondo i termini usati da Williams, un "chiaro percorso deliberativo" dal suo insieme motivazionale *esistente* all'attività del continuare a fumare, e non c'era un preciso percorso deliberativo dal suo insieme motivazionale esistente alla decisione di non fumare. Secondo il Modello Classico il suo era un caso di perfetta razionalità.

Io penso che questo caso riveli abbastanza chiaramente i limiti del Modello Classico, perché è un caso in cui qualcuno deve prendere una decisione corrente, e una decisione razionale richiede che si agisca in base a una ragione indipendente dai desideri. Perché, esattamente, il suo comportamento era irrazionale? Non penso che questo sia un caso difficile. L'irrazionalità deriva dal fatto che lo stesso sé che prende la decisione nel presente è lo stesso sé che morirà all'età di sessant'anni. Non è sufficiente dire che proprio in quel momento lei non aveva desideri circa i suoi desideri, e che dunque non aveva desideri riguardo al suo futuro. Il problema è che, in termini razionali, ella *dovrebbe* aver avuto desideri riguardo al suo futuro, perché il suo comportamento presente è tale che lei sta soddisfacendo e insieme distruggendo lo stesso sé. Notate che non sto sostenendo che "la gratificazione posticipata" sia sempre una scelta razionale. Mi sembra chiaro che possono esserci alcuni tipi di soddisfazioni, proprio ora, per le quali vale la pena rischiare la vita pur di raggiungerle. In tali casi si può costruire una ragione totale in cui è necessario mantenere un equilibrio tra la soddisfazione presente e il rischio della cessazione delle proprie speranze future. Ma questo non è un caso del genere. Nel nostro caso non c'era un soppesare la voglia di fumare ora e la non desiderabilità di morire in seguito. Il punto è che secondo il Modello Classico la non desiderabilità di morire in seguito non conta per nulla, perché non è rappresentato come parte dell'insieme motivazionale.

VI. CHE COS'È UNA RAGIONE PER UN'AZIONE?

La nostra domanda originale, che cos'è una ragione per l'azione, ora è stata trasformata: come abbiamo visto, una ragione per un'azione è una qualsiasi entità fattiva che è un elemento di un insieme che costituisce una ragione totale. Così l'obiettivo dell'analisi è il concetto di ragione totale. Che cos'è, allora, una ragione totale?

Una ragione totale per un'azione deve avere i seguenti componenti. Per prima cosa deve avere uno o più motivatori razionali. Che cosa rende razionale un motivatore? In

termini formali si può dire che un motivatore razionale deve essere un desiderio razionale oppure un qualche motivatore esterno razionale, come un obbligo, un impegno, un dovere, una richiesta o un bisogno. Per esempio il mio desiderio di pranzare e il mio bisogno di vitamine sono dei motivatori razionali. Ma il mio improvviso bisogno di addentare un pezzo di questo tavolo non è un motivatore razionale. Per funzionare in una decisione razionale il motivatore deve essere riconosciuto come tale dall'agente.

Secondo, ad eccezione di alcuni casi molto semplici in cui posso soddisfare il motivatore compiendo un'azione di base come sollevare il mio braccio, una ragione totale deve contenere un insieme di effettuatori e di costitutori. Queste entità fattive devono essere in relazione con i motivatori al fine di causare la soddisfazione del motivatore (questi sono gli effettuatori) oppure per cossituire esse stesse la soddisfazione del motivatore (questi sono i costitutori). La deliberazione razionale consiste dunque nel considerare i motivatori in base alla loro validità e in base ai conflitti tra i motivatori, e nel valutare gli effettuatori e i costitutori in modo tale da produrre la massima soddisfazione dei motivatori con il minor dispendio di altri motivatori nel soddisfare gli effettuatori e i costitutori. Per esporre questo punto in italiano corrente, per pensare razionalmente a che cosa fare bisogna individuare che cosa davvero si dovrebbe fare, e in seguito bisogna individuare il modo migliore per farlo senza frustrare una quantità di altre cose che si vogliono o che si debbono fare.

Alla luce di questa discussione, possiamo ora ritornare alla nostra domanda originale espressa nel paragrafo I e riformularla come segue:

Un agente razionale X considera correttamente un insieme di enunciato S , che consiste di enunciati individuali $s_1, s_2, s_3...$, come esprimente una ragione totale valida, per lui stesso o per lei stessa, per compiere un'azione del tipo A sse:

1. Ogni elemento di S , s_1, s_2 ecc. è vero e ritenuto vero da X .
2. S contiene l'enunciato di almeno un motivatore razionale e quel motivatore razionale è riconosciuto da X come motivatore razionale. I motivatori razionali, come abbiamo visto in precedenza, possono essere sia esterni che interni; essi possono, per esempio, essere desideri o obblighi, ma se l'obbligo deve funzionare internamente, deve essere riconosciuto come tale dall'agente.
3. X ritiene che S non enunci condizioni causalmente sufficienti per la realizzazione dell'azione A . Questo è il punto in cui compare il divario. Affinché X si impegni nel

prendere una decisione razionale, egli deve assumere di avere a disposizione una scelta genuina.

4. X considera alcuni degli enunciati di S come esprimenti degli effettuatori o dei costitutori (o entrambi) per i motivatori.
5. La valutazione razionale delle relazioni tra motivatori in competizione e le varie richieste degli effettuatori e dei costitutori, sono sufficienti a giustificare la scelta di A come una decisione razionale, considerato tutto, dato S .

Fino a questo punto questa caratterizzazione è puramente formale. Non abbiamo ancora detto che cosa rende razionale un motivatore, o come può essere che la razionalità riconoscitiva possa richiedere che un agente debba riconoscere un fatto esterno come motivatore, o quali sono le procedure per mezzo delle quali si suppone che noi dobbiamo valutare i diversi motivatori, costitutori e effettuatori, per arrivare a una decisione razionale. Prenderò in considerazione alcune di queste questioni nei capitoli seguenti. Tuttavia, faccio un avvertimento proprio ora: una teoria della razionalità non fornirà mai di per sé un algoritmo per prendere decisioni razionali. Una teoria della razionalità non fornirà un algoritmo per le decisioni razionali più di quanto una teoria della verità fornisca un procedimento per scoprire quali proposizioni sono vere. Una teoria della verità vi dice che cosa significa dire che una proposizione è vera e una teoria della razionalità vi dirà che cosa significa dire che un'azione era razionale.

P21

5.

ALCUNE CARATTERISTICHE SPECIALI DELLA RAGIONE PRATICA: UN ALTRUISMO FORTE COME REQUISITO LOGICO

I. *Le ragioni per l'azione*

Ho insistito molto sul fatto che nell'indagine sulla razionalità dovremmo concentrare la nostra attenzione sul ragionamento intendendolo come un'attività in cui effettivamente i sé si impegnano, piuttosto che concentrarci sulla razionalità intesa come un insieme astratto di proprietà logiche. Se lo facciamo, troveremo allora in ogni attività del ragionamento una collezione di fenomeni intenzionali e un sé che cerca di organizzarli in modo da produrre un altro stato intenzionale come prodotto finale. Nella ragione teorica il prodotto finale è una credenza o l'accettazione di una proposizione; nella ragione pratica è un'intenzione precedente o un'intenzione-in-azione. Una conseguenza dell'analisi dell'intenzionalità dell'azione che ho sviluppato nel capitolo 2 è che le azioni hanno contenuti intenzionali. Così non c'è nulla di misterioso nel fatto che le azioni possano essere il risultato di un processo di ragionamento. Proprio come la ragione teorica termina in una credenza o nell'accettazione di una proposizione, così la ragione pratica termina in un'intenzione precedente per agire o in un'azione effettiva (che ha il contenuto intenzionale di un'intenzione-in-azione). Spesso, ma non sempre, queste sono precedute dalla formazione di un desiderio secondario. Per esempio: io guardo fuori e giungo alla conclusione che sta per piovere. Dato il mio desiderio primario di rimanere asciutto e le altre mie credenze, io formo il desiderio secondario di portare un ombrello, l'intenzione precedente di portare il mio ombrello, ed esco di casa portando un ombrello. Ognuno degli ultimi tre passaggi, inclusa

l'azione stessa, ha un contenuto intenzionale motivato dai passaggi precedenti. Ho sentito persone sogghignare di fronte alla pretesa apparentemente bizzarra di Aristotele secondo cui un'azione può essere la conclusione di un "sillogismo pratico". Aristotele aveva ragione e coloro che sogghignavano avevano torto.

Ho sottolineato con enfasi il senso in cui la ragione teorica è un caso speciale della ragione pratica: decidere quali credenze accettare e quali rifiutare è un caso speciale del decidere che cosa fare. Sebbene sia la ragione teorica che quella pratica portino a un divario dove l'agente deve semplicemente agire, le ragioni per agire sono sotto molti aspetti differenti dalle ragioni per credere. Le ragioni per credere consentono una prova conclusiva in un modo che non è consentito dalle ragioni per agire. Questa è una conseguenza della differente direzione di adattamento. In questo paragrafo esplorerò alcune caratteristiche speciali delle ragioni per l'azione e le conseguenze che esse hanno per la ragione pratica. Che cosa c'è di speciale nelle ragioni per l'azione? Quali sono le differenze tra le ragioni per fare qualcosa e le ragioni per credere o accettare qualcosa? In entrambi i casi abbiamo un insieme di contenuti intenzionali con direzioni di adattamento verso l'alto e verso il basso. Si suppone che la direzione delle cose che si adattano verso il basso sia vera, così esse sono responsabili degli stati di cose nel mondo. Che tipo di direzione abbiamo per cose che si adattano verso l'alto e di che cosa esse sono responsabili? Nel caso della ragione teorica la risposta è relativamente semplice. Avere una credenza significa essere legati alla sua verità, così se sono legato alla ragione teorica sulla base delle mie credenze, io sono legato alla verità. Il legame ha una direzione di adattamento mondo-a-mente o verso l'alto, e il legame nei confronti della verità fornisce una ragione per accettare le proposizioni vere. Dire che qualcosa è vero implica che bisogna crederci. Per esprimerlo in maggior dettaglio: supponiamo che io voglia sapere se credere che p . Supponiamo che io abbia la prova conclusiva che p è vera. Dal momento che la credenza implica un legame con la verità e che il legame ha una direzione di adattamento verso l'alto, io devo credere (accettare, riconoscere o essere consapevole) che p .

Sia la ragione pratica che quella teorica sono soggette a vincoli razionali, ma le ragioni per l'azione hanno alcune caratteristiche speciali aggiuntive. Per prima cosa, le ragioni per agire hanno una specie di *status* di prima persona che le ragioni per credere non hanno. Le ragioni per credere si presentano caratteristicamente in forma di evidenza o di prova della verità della proposizione creduta, e la verità è impersonale. La verità è una

ragione per credere per tutti. Ma per quanto riguarda l'azione, perfino se la ragione è una ragione per tutti, le ragioni per l'azione devono ancora appellarsi a qualcosa di interno o in prima persona in un modo in cui le ragioni per credere non fanno. Una volta che si è stabilita la *verità* non esiste una questione ulteriore riguardo a ciò che dovrete credere, perché avere la credenza che *p* è vera significa già avere la credenza che *p*. Ma a causa della differenza nella direzione di adattamento, tra credenza e intenzione non c'è nulla di analogo alla verità per quanto riguarda le ragioni per l'azione. Nella ragione teorica le ragioni giuste vi portano a una credenza che è vera. Nella ragione pratica le ragioni giuste vi portano a un'intenzione che è... cosa? Non c'è un *x* tale che quell'intenzione sta a *x* come la verità sta alla credenza. Ognuno di noi ha una ragione per cercare l'auto-conservazione, la prosperità, l'autonomia e una quantità di altri obiettivi desiderabili. Ma nessuno di questi obiettivi sta all'azione come la verità sta alla credenza, poiché in ogni caso l'obiettivo deve essere rappresentato dai contenuti intenzionali dell'agente come un obiettivo separato. Nel caso della credenza, l'obiettivo della verità è costruito all'interno della credenza. Nessun obiettivo del genere è costruito all'interno delle ragioni per agire, delle intenzioni precedenti o delle intenzioni-in-azione.

In secondo luogo, le ragioni per agire hanno una relazione speciale nei confronti del tempo che è diversa da quella delle ragioni per credere. Le ragioni per agire sono sempre rivolte al futuro. E questo è vero perfino in quei casi in cui forniamo le ragioni per le quali un agente ha agito in un determinato modo nel passato. Una ragione presente per agire è sempre una ragione per un sé per compiere un'azione sia *ora* che *dopo*. Una ragione per l'azione passata era una ragione nel passato perché un sé compisse quell'azione in quel momento o più tardi.

Il terzo punto è legato a questi due. Le ragioni per un'azione devono essere capaci di motivare un'azione. Se è data una ragione del perché un'azione è stata compiuta nel passato, allora la ragione deve aver funzionato causalmente nell'esecuzione dell'azione, poiché essa deve essere stata la ragione *sulla base della quale l'agente ha agito*. Se la ragione riguarda un'azione futura, allora deve esserci una ragione *sulla base della quale l'agente può agire*. Ma dire questo significa dire che la ragione è effettiva sia realmente che potenzialmente, poiché la nozione di agire sulla base di una ragione, come abbiamo visto, è la nozione di rendere la ragione effettiva all'interno dell'esecuzione dell'azione. Nel capitolo precedente

ho richiamato l'attenzione sulla caratteristica motivazionale delle ragioni per ipotizzare che ogni ragione totale deve contenere almeno un motivatore.

Che tipo di fattivi possono essere i motivatori? La risposta data dal modello classico è brutalmente semplice: tutti i motivatori sono desideri, dove “desiderio” è concepito in maniera generale per includere cose come le mete, gli scopi e gli obiettivi dell'agente. La ragione è e deve essere schiava delle passioni. Autori recenti sono in un qualche modo vaghi riguardo a ciò che l'elenco delle entità motivazionali dovrebbe includere, ed essi parlano in modo generale di “pro-attitudini” (un termine inventato, credo, da Patrick Nowell-Smith)¹ e di “complesso, o insieme, motivazionale soggettivo” (Williams)², ma l'idea generale è abbastanza chiara. Senza qualche tipo di stato psicologico interno come i desideri, il processo del ragionamento non potrebbe mai produrre un'azione. Gli scimpanzé di Köhler ne sono il modello. Senza desiderio non si sarebbero mai staccati da terra.

Perché i teorici classici confidano così tanto in questo modello? Ebbene, la sua semplicità è affascinante e rende le sue caratteristiche formalizzabili in modo elegante nella teoria della decisione. Ma vi sono anche importanti motivi filosofici per sostenerlo. Per prima cosa, nella vita reale, vi sono molti casi di questo tipo. I casi più semplici sono quelli in cui la ragione è semplicemente un desiderio di un certo tipo. “Perché stai bevendo dell'acqua?” Perché ho sete. Un altro tipo di caso è quando c'è qualche fatto che l'agente crede porterà alla soddisfazione del suo desiderio. “Perché stai bevendo dell'acqua?” Perché curerà il mio mal di testa. Ecco l'intera storia: io voglio curare il mio mal di testa, credo che bere dell'acqua curerà il mio mal di testa, perciò voglio bere dell'acqua. In un caso del genere il desiderio di bere dell'acqua è esso stesso un desiderio motivato, motivato da un altro desiderio assieme ad una credenza che riguarda il come soddisfare quel desiderio.

Un altro argomento a favore del modello classico è che nella struttura della deliberazione effettiva la conclusione deve essere un qualche stato intenzionale del tipo dei desideri, come un desiderio secondario, un'intenzione precedente, o un'intenzione-in-azione. E da dove potrebbe provenire razionalmente quello stato se non da un desiderio precedente? Senza un desiderio o una pro-attitudine come punto di partenza, sembra non ci sia un modo in cui la deliberazione potrebbe razionalmente finire in un desiderio o in uno stato intenzionale del tipo dei desideri.

¹ Patrick Nowell-Smith, *Ethics*, London, Penguin Books, 1954, p. 112, Patrick Nowell-Smith, *Etica*, La Nuova Italia, Firenze 1974, p. 123.

L'obiezione ovvia alla pretesa del modello classico secondo la quale solo i desideri possono fornire delle motivazioni è che ci sono molte ragioni per l'azione effettive dal punto di vista motivazionale, come gli obblighi, che non sono dei desideri. "Perché stai bevendo dell'acqua?" Perché sono obbligato a farlo. L'ho promesso alla mia sposa.

A tutti questi esempi il modello classico fornisce la stessa risposta. Il vostro obbligo, per esempio, è una ragione per l'azione semplicemente perché voi *desiderate* assolvere i vostri obblighi. Uno dei punti centrali nella discussione tra la mia posizione e il modello classico è esattamente su questa questione. Secondo la mia prospettiva l'obbligo è – o per lo meno potrebbe essere – la ragione per un desiderio effettivo (per esempio, il desiderio sulla base del quale l'agente agisce), piuttosto che un desiderio precedente che funge da ragione perché l'obbligo sia effettivo. Ritorno su questo punto nel prossimo capitolo.

Una quarta caratteristica delle ragioni per agire è che se la ragione è intesa come la ragione per il compimento di un'azione libera, essa non può essere ritenuta dall'agente come causalmente sufficiente. Se egli pensa di essere veramente costretto, allora non può pensare di agire liberamente sulla base di una ragione. Nel caso delle azioni umane, a causa del divario, la ragione può essere buona o adeguata senza che vengano fornite condizioni per l'azione causalmente sufficienti. E, più importante dal punto di vista dell'agente, la ragione non deve essere vista come causalmente sufficiente. Come ho fatto notare nei capitoli precedenti, l'applicabilità del concetto di razionalità nel prendere decisioni presuppone la libera scelta. Per gli agenti razionali, la libera scelta è in effetti sia necessaria che sufficiente per l'applicabilità della razionalità. La libera scelta implica che l'azione sia razionalmente valutabile, e la possibilità di essere valutata dal punto di vista razionale implica la libera scelta. Potrebbe sembrare che vi siano moltissimi controesempi nei confronti di questa posizione. "Che dire del tossicodipendente che non riesce a frenarsi, ma ciononostante è capace di razionalità nel selezionare dei mezzi razionali piuttosto che irrazionali, per soddisfare la sua voglia irresistibile?" Ma perfino questo caso sostiene il punto generale, poiché noi stiamo tacitamente supponendo che il tossicodipendente abbia a disposizione una scelta di mezzi per soddisfare il suo desiderio irrefrenabile. Cioè, se consideriamo l'agente come un soggetto che agisce razionalmente, significa che supponiamo che in quell'ambito egli stia compiendo delle scelte libere, anche se il progetto generale di soddisfare la sua dipendenza non è una questione di libera scelta e quindi cade al di fuori dell'ambito della

² "External and Internal Reasons", ristampato in *Moral Luck*, Cambridge, Cambridge University Press, 1981, pp.

razionalità. Il divario è una caratteristica che riguarda sia il ragionare su che cosa credere che il ragionare su che cosa fare. Ma esso gioca un ruolo speciale nel ragionare su che cosa fare, come ho cercato di descrivere.

Per riassumere: in aggiunta ai due vincoli generali della *razionalità*, (assieme alla giustificazione), e al *divario*, che si applicano alle ragioni per credere così come alle ragioni per fare, ci sono almeno tre caratteristiche speciali che riguardano le ragioni per l'azione. Esse sono, in un senso speciale in *prima persona*, esse sono essenzialmente *rivolte al futuro* e sono essenzialmente *motivazionali* nel senso che devono essere capaci di motivare un'azione. Per usare delle parole importanti, chiamiamo queste cinque le condizioni della Razionalità, Libertà, Soggettività, Temporalità e Causazione.

Perché queste condizioni dovrebbero stare assieme in questo modo? Perché ci sono queste connessioni? Ad un certo livello non credo che questa sia una domanda difficile. La razionalità è un fenomeno biologico. La razionalità in azione è quella caratteristica che permette agli organismi, con un cervello sufficientemente grande e complesso, di avere dei sé coscienti, di coordinare i loro contenuti intenzionali, in modo da produrre azioni migliori di quanto non farebbero con un comportamento casuale, con l'istinto, col tropismo o agendo sulla base di un impulso. Per ottenere dei vantaggi biologici dal comportamento razionale gli animali devono avere dei propri motivi coscienti (Soggettività), alcuni di questi devono essere rivolti al futuro (Temporalità), essi devono essere capaci di motivare il comportamento reale nella forma di movimenti corporei (Causazione) e devono farlo sotto il presupposto della libertà che opera nel divario (Libertà). La "ragione pratica" è il nome di tale capacità di coordinazione. Infatti queste caratteristiche non sono logicamente indipendenti: le prime due caratteristiche, la Soggettività e la Temporalità, sono una conseguenza della terza caratteristica, la Causazione motivazionale. Un motivo deve essere il motivo di qualcuno (Soggettività) per agire ora o nel futuro (Temporalità).

La connessione tra la razionalità e il divario della libertà è questa: *la razionalità si applica solamente dove c'è una scelta libera, poiché la razionalità deve essere capace di produrre una differenza*. Se le mie azioni sono davvero completamente *causate* dalle mie credenze e desideri, in modo che io non posso controllarmi, allora non ho scelta e la razionalità non può produrre nessuna differenza nel mio comportamento. Se sono sotto la costrizione di condizioni causalmente sufficienti, non c'è spazio per la deliberazione e la

mia azione cade al di fuori dell'ambito della valutazione razionale. Inoltre una richiesta di giustificazione ha senso solo nei casi in cui si aprissero per l'agente delle possibilità alternative.

II. Costruire un animale razionale

Per illustrare il ruolo speciale e la caratteristica della ragione pratica, mi piacerebbe presentare il seguente esperimento mentale. Immaginate di progettare e costruire un robot destinato a diventare un "animale razionale". Il punto essenziale dell'esperimento mentale è illustrare le relazioni logiche tra certe caratteristiche cruciali dell'esistenza umana. Qualunque cosa noi siamo, siamo dei prodotti, almeno in senso metaforico, di un certo tipo di progettazione. Non credo si sia trattato della progettazione divina della visione creazionista, ma piuttosto, per quanto ne sappiamo, si è trattato della progettazione non intenzionale, metaforica, del tipo "come se" dei processi evolutivi. In un modo o nell'altro, noi siamo il risultato di un certo insieme di processi che sono stati guidati da particolari di necessità di progettazione. Dato che siamo il prodotto di una progettazione, anche se solo di quella del tipo "come se", il senso del domandare come si possa progettare un essere razionale è dato dal riuscire a vedere che cosa dobbiamo inserire nel nostro progetto per vedere che cosa potremmo ottenere come risultato di quanto è stato inserito.

Che cosa si richiede come effettiva caratteristica di un progetto, e che cosa si ritiene invece marginale? (a proposito, molte delle questioni della storia della filosofia sono contenute in questa domanda). Poiché la razionalità non è una facoltà separata o un modulo, ma piuttosto una caratteristica interna ad altre capacità cognitive e volizionali, credo scopriremo di dover inserire gran parte, sebbene non tutte, delle facoltà mentali umane per ottenere una "macchina" capace di razionalità.

La prima caratteristica che bisogna inserire nel robot è la coscienza. E' necessario costruire un cervello artificiale che abbia il potere del cervello umano di causare e sostenere stati soggettivi di consapevolezza e sensibilità interni, qualitativi e unificati. Senza la coscienza non si può entrare in alcun modo nel gioco della razionalità. Ma la coscienza percettiva passiva non è sufficiente. E' necessaria la coscienza attiva dell'agire. Bisogna cioè costruire un essere che sia coscientemente capace di intraprendere delle azioni. Ma per

fare questo il robot deve avere desideri e intenzioni. Questo perché deve essere in grado di voler fare le cose che tenta di fare. Così, come requisito minimo, dobbiamo avere una macchina capace di percezione, azione e desiderio. Inoltre, se queste azioni devono essere azioni razionali, il robot deve essere capace di deliberazione. Questo requisito è molto più importante di quanto potrebbe sembrare a prima vista. Non vedo come un robot potrebbe agire deliberatamente senza disporre di una parte molto ampia dell'apparato umano e animale dell'intenzionalità. Per prima cosa, deve esserci la capacità di immagazzinare le informazioni in forma di ricordi, e questa capacità di ricordare sarebbe la fonte delle credenze. Secondo, esso dovrebbe possedere la capacità di coordinare sia le cose che hanno direzione di adattamento verso il basso (credenze, percezioni ecc.) che quelle che hanno direzione di adattamento verso l'alto (desideri, inclinazioni ecc.) in una corrente di pensieri coscienti. Non è dunque sufficiente avere delle percezioni, dei ricordi, dei desideri e delle intenzioni; il robot deve anche essere capace di mettere in funzione tutto questo apparato in una sequenza cosciente di pensieri deliberativi. Deve essere capace di pensare che poiché si trova nella tale situazione, e desidera la tal cosa, dovrebbe compiere questa azione e non quest'altra, anche se può pensare questi pensieri soltanto senza parole. Per possedere tutto questo apparato intenzionale esso deve avere ciò che ho chiamato (nel capitolo 2) lo Sfondo, l'insieme di capacità pre-intenzionali che gli consentono di interpretare e di applicare i suoi stati intenzionali. Infine, il robot deve essere tale che il flusso di pensieri sia capace di terminare in decisioni e in azioni conseguenti.

Gli elementi che abbiamo dovuto aggiungere al robot dopo avergli dato una coscienza sono state dunque piuttosto sostanziali: il robot deve avere fenomeni percettivi coscienti, fenomeni conativi coscienti (desideri), e fenomeni volitivi coscienti (intenzioni precedenti e intenzioni-in-azione) e deve avere la capacità di far sì che la sua deliberazione cosciente si manifesti in decisioni e in azioni, con tutto l'apparato che tale processo deliberativo implica. Nella mia descrizione del caso, è già inserita l'esperienza del divario nel robot. E poiché possiede tutte queste caratteristiche, come ho osservato nel capitolo 3, esso possiede già un sé, nel senso che ho specificato. Il sé, come lo intendo io, giunge spontaneamente quando si ha un essere intenzionale cosciente capace di compiere azioni libere sulla base di ragioni. Ora sorge immediatamente una domanda cruciale. Una volta che il robot abbia tutto questo, possiede già il meccanismo necessario per prendere decisioni razionali con la varietà propria dell'essere umano? Bene, non proprio. Fino ad ora noi non abbiamo costruito un robot

umanoide ma, si potrebbe dire, uno scimpanzé artificiale. Per giungere ai poteri propri della capacità umana di prendere decisioni abbiamo bisogno di inserire certe altre caratteristiche.

Una volta che si abbiano stati mentali e processi sia coscienti che non coscienti, con direzione di adattamento sia verso il basso (percezioni, memorie, credenze ecc.) che verso l'alto (desideri, inclinazioni, intenzioni ecc.), e si abbia la capacità di coordinare tutto questo nel flusso di pensiero cosciente che termina nel prendere decisioni, il successivo elemento centrale da costruire nel robot è, senza dubbio, il linguaggio. È importante dire esattamente quali caratteristiche del linguaggio sarebbero richieste da un agente razionale. Un animale non ha bisogno di alcun linguaggio per avere semplici stati intenzionali come la fame e la sete, e non deve possedere il linguaggio nemmeno per prendere decisioni semplici, né in verità ha bisogno del linguaggio per compiere un semplice ragionamento strumentale come quello degli scimpanzé di Köhler. Ma per una razionalità completa certe caratteristiche specifiche del linguaggio sono essenziali. Non tutte le caratteristiche dei linguaggi naturali umani sono essenziali per la razionalità. Per esempio, i processi razionali del pensiero non richiedono parole per i colori, per le voci passive o per gli articoli determinativi. Ma la razionalità completamente umana ha bisogno di certi meccanismi linguistici essenziali. Per prima cosa, il nostro robot deve possedere le forme di base dell'atto linguistico che connette il linguaggio alla realtà sia con direzione di adattamento parola-a-mondo che con direzione di adattamento mondo-a-parola. Deve, come minimo, avere la capacità di rappresentare come stanno le cose nel mondo (assertivi), così come la capacità di rappresentare in che modo esso tenta di portare gli altri ad agire nel mondo (direttivi) e come esso si impegna ad agire nel mondo (commissivi). Inoltre, deve avere la capacità di comunicare tutto questo ad altri possessori di linguaggio. Il linguaggio comprende sia il pensare che il parlare, ma quando parliamo, dobbiamo disporre di un linguaggio che sia pubblico, che permetta al robot di comunicare con altri. Poiché stiamo costruendo questo robot a nostra immagine, per così dire, lo costruiremo con la capacità di comunicare con noi. Inoltre, mi sembra che il robot debba disporre di alcuni dispositivi per rappresentare le relazioni temporali. Per essere in grado di progettare il futuro, che è caratteristico della ragione pratica, deve essere capace di rappresentare il futuro e la sua relazione con il presente e con il passato. Di che cos'altro avrebbe bisogno? Ebbene, mi sembra che dovrebbe essere in grado di articolare le relazioni logiche. Non deve per forza possedere il nostro vocabolario logico, ma dovrebbe essere in qualche modo capace di contraddistinguere la negazione, la congiunzione, l'implicazione e

la disgiunzione. Penso inoltre che dovrebbe possedere qualche insieme, se pure minimale, di termini metalinguistici per apprezzare successi e fallimenti nel realizzare la direzione di adattamento e la coerenza logica. Così esso ha bisogno di qualcosa nel campo che include “vero” e “falso”, “valido” e “non valido”, “accurato” e “non accurato”, “rilevante” e “irrilevante”. Ora che gli abbiamo dato quel tanto di linguaggio che basta, possiamo anche dargli un nome. Chiamiamolo “la Creatura”.

Nel costruire tutto questo apparato rappresentativo, che comprende rappresentazioni mentali e linguistiche, dovremo fornire alla Creatura l'apparato necessario ad applicare queste rappresentazioni alle situazioni concrete e ad interpretare le rappresentazioni che riceve da altre fonti. Queste abilità, le abilità di applicare e interpretare le rappresentazioni, costituiscono ciò che ho chiamato lo Sfondo.

Questo è il punto cruciale dell'esperimento mentale: una volta che la Creatura ha tutto questo, essa possiede già l'apparato essenziale delle caratteristiche peculiarmente umane dei processi mentali razionali e del comportamento razionale. Essa ha una forma di razionalità che va ben oltre quella dello scimpanzé razionale del quale abbiamo discusso nel capitolo 1. Specificamente, una volta che la Creatura ha l'abilità di compiere degli atti linguistici, essa ha il potenziale per le ragioni per l'azione indipendenti dal desiderio; essa ha inevitabilmente il requisito delle ragioni per l'azione indipendenti dal desiderio, poiché quasi ogni atto linguistico implica un *impegno* di qualche tipo. Gli esempi noti sono quelli degli atti linguistici come il promettere, dove il parlante è impegnato a portare a compimento un futuro corso di azione, ma l'asserire impegna il parlante nei confronti della verità della proposizione asserita e gli ordini impegnano il parlante a credere che la persona alla quale egli impartisce l'ordine sia capace di eseguirlo, lo impegnano a desiderare che ella lo faccia, e a permettere che chi lo ascolta possa farlo. In breve, ciò a cui la gente pensa come elemento che caratterizza del promettere, vale a dire l'impegno o l'obbligo, in realtà riguarda tutti gli atti linguistici. Le sole eccezioni cui riesco a pensare sarebbero semplici espressioni come “Accidenti!” “Maledizione!” o “Urrà” e perfino queste impegnano il parlante ad avere certe attitudini.

La caratteristica bizzarra della nostra tradizione intellettuale, secondo la quale nessun insieme di enunciati veri che descrivano come le cose sono nel mondo può mai implicare logicamente un enunciato che dica come queste cose dovrebbero essere, è che l'effettiva terminologia con cui questa tesi viene enunciata contraddice la tesi stessa. Così, per

esempio, dire che qualcosa è vero significa già dire che si dovrebbe crederlo, che, a parità di altre cose, non si dovrebbe negarlo. La nozione di inferenza valida è tale che se p può essere validamente inferito da q , allora chiunque asserisca p non dovrebbe negare q ; ovvero, chiunque sia impegnato in p dovrebbe riconoscere di essere impegnato in q .

Il punto cruciale dell'esperimento mentale può essere posto anche in questo modo: qualora si abbia l'apparato della coscienza e dell'intenzionalità, e un linguaggio abbastanza ricco da poter realizzare i diversi tipi di atti linguistici ed esprimere le diverse relazioni logiche e temporali, allora si possiede già l'apparato necessario per la razionalità. La razionalità non è un modulo o una facoltà ulteriore (*extra*). Essa è già insita nell'apparato che abbiamo descritto. Inoltre, qualcosa di più ricco della razionalità strumentale o del tipo mezzi-fini è già compreso nell'apparato che abbiamo descritto, poiché abbiamo il potenziale, ovvero il requisito, di ragioni per l'azione indipendenti dal desiderio o esterne.

Abbiamo incluso nella creatura le esperienze del divario. Ma le abbiamo dato il vero libero arbitrio o soltanto l'illusione del libero arbitrio? Ci sono almeno due differenti possibilità. Per prima cosa potremmo illudere la povera Creatura rendendo i suoi meccanismi totalmente deterministici. In questo modo essa avrebbe l'illusione del libero arbitrio, poiché esperirebbe il divario, ma di fatto il suo comportamento sarebbe interamente programmato secondo meccanismi totalmente deterministici. Un'altra possibilità piuttosto diversa è che alla sua esperienza cosciente del prendere una decisione, all'interno del divario, corrisponda un elemento indeterministico nella implementazione dell'*hardware* che è sviluppato nel tempo dal fatto di prendere una decisione a livello cosciente. Esplorerò entrambe queste possibilità, per quanto riguarda gli esseri umani, nel capitolo 9.

III . Egoismo e altruismo nella Creatura

Bene, che cosa possiamo dire dell'argomento preferito dai filosofi morali, l'egoismo e l'altruismo? Come si relazionano con il nostro robot? Non abbiamo ancora esplicitamente inserito né l'altruismo né l'egoismo nella Creatura. Nella nostra cultura stimiamo l'egoismo e l'auto-interesse come non problematici e ci riferiamo all'altruismo e alla generosità come argomenti che richiedono una spiegazione speciale. In un certo senso questo è corretto, ma in un altro senso è sbagliato. È giusto supporre che la Creatura preferirà la soddisfazione dei

suoi desideri alle sue frustrazioni, e preferirà l'alleviarsi dei suoi dolori piuttosto che la loro intensificazione. A parità di condizioni, questo fa parte dell'avere un desiderio o un dolore. E per quanto riguarda i suoi desideri ecc. questo sembra egoismo. Ma in un altro senso è sbagliato pensare all'egoismo come non problematico, poiché la soddisfazione dei desideri non ci dice qual è il *contenuto* dei desideri, e finora non abbiamo detto nulla riguardo al contenuto dei desideri della Creatura. Potrebbe essere che la Creatura consideri i desideri altruistici naturali quanto i desideri egoistici. Per quanto abbiamo detto finora, la Creatura potrebbe preferire la prosperità degli altri alla propria.

Aggiungiamo quindi un altro componente alla nostra Creatura. Supponiamo di programmarla per cercare ciò che chiamerò vagamente "auto-interesse". Inseriamo nella nostra Creatura la preferenza per la sopravvivenza piuttosto che per l'estinzione e una preferenza per il suo auto-interesse piuttosto che per ciò che non è suo proprio interesse, cioè supponiamo che la Creatura non desideri essere ferita, danneggiata, ammalata, defraudata o morta. Una volta che la Creatura ha un sé e un auto-interesse, se ha anche una concezione del tempo, come abbiamo stabilito, allora sarà capace di progettare la sua conseguente sopravvivenza e prosperità. Cioè, se il sé ha degli interessi, se il sé persiste attraverso il tempo e se il sé è l'agente che esercita la razionalità, allora risulterà razionale per il sé elaborare dei piani al tempo presente per assicurare i suoi interessi nel futuro, anche se non avesse un desiderio presente di fare delle cose necessarie per assicurare i suoi interessi nel futuro. Abbiamo ora due forme di ragioni per l'azione indipendenti dai desideri o esterne. Per dirla in maniera grossolana, ci sono impegni, che normalmente si prendono verso gli altri, ma tali impegni possono essere presi anche verso sé stessi; e ci sono ragioni prudenti.

L'auto-interesse razionale nel nostro robot illuminato non giunge spontaneamente, ma non richiede nemmeno un grande investimento tecnologico, se non il minimo necessario per la coscienza, l'intenzionalità e il linguaggio. Se la Creatura ha dei bisogni e degli interessi ed ha la capacità di riconoscere questi bisogni e questi interessi, e se possiede un sé e una consapevolezza del fatto che il suo sé si estende nel futuro, non vi sarà un elemento ulteriore a fornirle una motivazione per agire ora in modo da fare attenzione ai suoi interessi nel futuro.

Ed ora veniamo ad una domanda cruciale: la Creatura ha una base razionale per preoccuparsi degli interessi degli altri? Qual è la relazione tra l'auto-interesse che abbiamo

inserito e l'altruismo che abbiamo tralasciato? L'approccio tradizionale a questa domanda da parte dei filosofi morali è quello di cercare di far derivare l'altruismo dall'egoismo. Ci sono, se ho capito bene ciò che intendono, almeno tre modi per fare questo. Per prima cosa immaginiamo di farlo semplicemente come un compito di ingegneria. Inseriamo l'altruismo nella nostra Creatura proprio come abbiamo inserito l'egoismo. Questo è un modo per interpretare il lavoro dei socio-biologi. L'idea è che noi siamo geneticamente inclini ad alcune forme di altruismo e si suppone che siamo capaci di spiegare le basi genetiche dell'altruismo tramite cose come la selezione dei gruppi o la selezione parentale. L'altruismo è semplicemente un'inclinazione naturale ed esso può essere effettivo come ogni altra ragione interna. La nostra Creatura ha semplicemente un'inclinazione a fare attenzione agli interessi degli altri. In secondo luogo, e questo è più interessante, è stato compiuto uno sforzo da parte di Thomas Nagel³ per dimostrare la somiglianza formale tra le ragioni prudenziali e le ragioni altruistiche. Considerare gli interessi degli altri è razionalmente fondato tanto quanto il considerare i propri interessi futuri. Terza e ultima cosa, è stato compiuto uno sforzo all'interno della tradizione kantiana, principalmente da Christine Korsgaard⁴, per derivare l'altruismo dall'autonomia. Se, a causa della mia autonomia o libertà, devo volere le mie azioni, e se il volere è soggetto a vincoli di generalità tali che mi viene razionalmente richiesto che ogni cosa che voglio dovrei essere in grado di volerla come se fosse una legge universale, allora mi verrà razionalmente richiesto di trattare le altre persone come miei pari nel regno morale a causa delle leggi universali che applicherò sia a me che agli altri.

C'è qualcosa di corretto in ognuno di questi tre approcci, ma anche qualcosa di insoddisfacente. Se io provo semplicemente un'inclinazione all'altruismo allora risulterà troppo fragile formare le basi per la ragione pratica in relazione all'altruismo. L'inclinazione all'altruismo non ha una speciale forza vincolante. Spesso non ci si accorge di queste inclinazioni e molte persone avvertono delle contro-inclinazioni, come l'inclinazione al sadismo, alla crudeltà o all'indifferenza. Secondo questa prospettiva, l'altruismo sarebbe semplicemente un'inclinazione tra le altre. Che cosa c'è di speciale nell'inclinazione ad aiutare gli altri? Ritorniamo all'analogia tra prudenza e altruismo fatta da Nagel. Il punto che mi sembra vero è il seguente: una volta che ho la coscienza e il sé, e che sono capace di

³ *The Possibility of Altruism*, Princeton, Princeton University Press, 1970, tr. it. *La possibilità dell'altruismo*, tr. it. Il Mulino, Bologna 1994.

⁴ *The Source of Normativity*, Cambridge, Cambridge University Press, 1996.

usare il linguaggio, sono già legato all'esistenza di altre coscienze e altri sé alla pari con me stesso. In che modo, esattamente? Il fatto che esista una cosa come il mio sé cosciente ha senso per me solo se esso è differente dalle altre cose dell'universo. Se c'è un io allora ci deve essere un non-io. E se le entità che costituiscono il non-io nell'universo includono entità con le quali io comunico nel compiere degli atti linguistici, allora devo presupporre che alcuni dei non-io nell'universo debbano essere degli agenti coscienti con un sé, proprio come me. Quindi io sono un sé tra gli altri. Ma la domanda ancora rimane: perché dovrei preoccuparmi degli altri? C'è, per la verità, una somiglianza formale tra il preoccuparsi del mio sé futuro e il preoccuparsi di un altro sé: in entrambi i casi devo considerare gli interessi di entità che non sono presenti alla mia coscienza proprio nel momento e nel luogo in cui sto prendendo le decisioni. Ma c'è un'asimmetria drastica: nel ragionamento prudenziale il sé di cui mi preoccupo sono io stesso. Cioè il sé che prende le decisioni e compie le azioni è identico al beneficiario delle decisioni e delle azioni. Per quel che riguarda il ragionamento altruistico, questa identità viene perduta. Non sto tentando in questa sede di rendere piena giustizia al sottile argomento di Nagel. Sto semplicemente sollevando una difficoltà che trovo in esso, prima di proseguire nel discutere un altro argomento che conduce alla stessa conclusione ed infine presentare il mio.

Passiamo allora ad esaminare la spiegazione kantiana della Korsgaard sul come l'autonomia generi l'universalità e su come l'universalità generi l'altruismo. La sua soluzione viene presentata come un'interpretazione delle posizioni di Kant ed ecco come si sviluppa: Kant sostiene che (1) dobbiamo agire sotto il presupposto del nostro libero arbitrio. Egli poi prosegue sostenendo che (2) il libero arbitrio, ammesso che ci sia una volontà, deve essere determinato in accordo con una legge. Dal momento che, perciò, (3) il libero arbitrio deve essere determinato in base alla sua *stessa* legge (da 1), risulta che (4) l'imperativo categorico è una legge del libero arbitrio.⁵ Il passo dubbio qui è il secondo. Perché l'esercizio del mio libero arbitrio dovrebbe mai richiedere una legge? Perché non posso liberamente decidere che cosa fare e basta? Certamente fino ad ora non è stato presentato nessun argomento sul perché debba esserci una legge affinché io operi delle decisioni razionali libere.

Per rispondere a questa obiezione la Korsgaard traccia un'analogia con la causazione. Ella sostiene che la causazione ha due componenti: la nozione del far sì che qualcosa accada

⁵ Korsgaard, *Sources of Normativity*, pp. 221-222.

e la nozione di legge. Noi richiediamo la seconda componente, la legge, perché non potremmo propriamente *identificare* un caso di qualcosa che fa sì che qualcos'altro accada senza assumerlo sotto una legge causale. Ovvero, ella pensa che la regolarità sia necessaria per l'identificazione della causazione. E sostiene poi che la causazione del volere è esattamente analoga alla causazione in generale: questo perché se agisco sulla base del mio libero arbitrio, allora sono la causa delle mie azioni. Ma se le cose stanno così, io devo essere capace di distinguere tra *me stesso* che causa l'azione e qualche *desiderio* o *impulso* che si trova in me che causa il movimento del mio corpo. Devo vedere me stesso come qualcosa di distinto dai miei desideri e dai miei impulsi di primo ordine. Ma se le cose stanno così, affinché le azioni siano realmente le mie azioni, cioè affinché esse provengano da me stesso piuttosto che essere espressioni dei miei desideri di primo ordine, io devo agire in base ad alcuni principi universali. Così, la legge che io creo per me stesso è esattamente analoga alle leggi della causazione. Noi non potremmo *identificare* le azioni come azioni di un sé se non fossero compiute in base ad un principio universale. Per poter veramente sostenere che le azioni sono mie, deve risultare che io sono un agente capace di attribuire delle leggi. Infatti è solamente perché imponiamo dei principi di volontà di carattere universale alle nostre decisioni che possiamo dire di avere un sé a tutti gli effetti. Il sé è costituito da queste decisioni di carattere universale. Per la Korsgaard, la frase chiave credo sia la seguente: “Se *tutte* le mie decisioni fossero particolari e anomale non ci sarebbe una differenza identificabile tra il *mio agire* e un *insieme di impulsi di primo ordine che sono causalmente effettivi all'interno o attraverso il mio corpo*. E quindi non ci sarebbe nessun sé – nessuna mente – nessun io – che sarebbe colui che compie l'azione” (p. 228).

Ritengo che questo argomento non funzioni. La nozione basilare di causazione è, infatti, la nozione del far sì che qualcosa accada. Ed è vero che per *identificare* tali accadimenti noi dobbiamo presupporre una regolarità. Ma questo è un requisito *epistemico*, non un requisito *ontologico*, riguardo la effettiva esistenza della causazione. Non c'è alcuna auto-contraddizione nell'immaginare delle cause che accadono senza fare riferimento a delle regolarità universali. Potremmo non essere capaci di stabilire con certezza che un certo particolare evento è stato davvero la causa di un certo altro evento, a meno che l'esperimento non sia ripetibile, a meno che non possiamo esaminare il caso individuale osservando se presenta una regolarità. Ma questa è una questione che riguarda lo scoprire con certezza; non è una questione che riguarda la reale esistenza della relazione tramite cui

una cosa ha fatto sì che un'altra cosa accadesse. Gli esempi della vita reale rendono chiara la distinzione tra causazione e regolarità. Quando, per esempio, indaghiamo le cause della prima Guerra Mondiale, noi cerchiamo di spiegare perché essa ebbe luogo. Non cerchiamo delle regolarità universali. Per intraprendere la ricerca dobbiamo presupporre uno Sfondo di alcuni gradi di regolarità, e senza la possibilità di condizioni causalmente sufficienti e di esperimenti ripetibili non potremmo mai essere completamente sicuri della nostra risposta. Ma il requisito di una regolarità è un requisito di carattere epistemico per l'*identificazione* delle cause; non è un requisito ontologico sull'esistenza reale della relazione tramite la quale un evento fa sì che ne accada un altro.

In effetti, il requisito della regolarità è un requisito epistemico che riguarda qualsiasi nozione che abbia un'applicazione nel mondo reale. Per identificare qualcosa come una sedia, un tavolo, una montagna o un albero, dobbiamo presupporre qualche tipo di regolarità presente nelle sue caratteristiche o nei suoi usi. La regolarità è essenziale per l'identificazione di un oggetto come una sedia, ma non dovremmo, su questa base, dire che la nozione di sedia contiene realmente due componenti, ovvero un oggetto che serve a far sedere la gente e un principio regolare. Piuttosto dovremmo dire che la sedia è un oggetto che la gente usa per sedersi e, come altre nozioni che si riferiscono agli oggetti, alle cause ecc., il concetto di sedia richiede una presupposizione di Sfondo di regolarità.

Se estendiamo la relazione di regolarità alla causazione nell'ambito degli esseri umani, possiamo dire che dal punto di vista della terza persona essa è in effetti un requisito epistemico del mio *riconoscere* le decisioni di qualcuno come decisioni considerate veramente sue, in quanto opposte al suo comportamento capriccioso e stravagante, e aventi un qualche tipo di ordine e di regolarità. Ma da questo non segue che, per *essere* le sue decisioni, esse debbano derivare da una legge universale che egli ha prodotto per se stesso. Questo equivale a dire che il passaggio che ho citato presenta una falsa dicotomia tra l'agire sulla base di un impulso, che si suppone non essere libero, e l'agire secondo una legge universale, che è libera. Ma l'agire sulla base di un impulso può essere libero come l'agire secondo una legge universale. La Korsgaard sostiene che non ci sarebbero differenze *identificabili* tra un atto non libero e un atto capriccioso, se tutti gli atti di una persona fossero capricciosi. Ma se questo punto è vero, è tuttavia soltanto un punto di vista epistemico in terza persona. Dall'esterno, qualcuno che mi sta guardando potrebbe non essere capace di dire quali delle mie azioni erano veramente libere, se io ho sempre agito

sulla base di un impulso. Ma dall'interno, dal punto di vista della prima persona, l'agire sulla base di un impulso può essere tanto libero quanto l'agire sulla base di una sobria riflessione. Alcune persone molto prudenti si trattengono dall'agire secondo l'impulso, mentre gli spiriti liberi spesso fanno sì che siano i loro impulsi a guidarli. L'esperienza del divario può essere la stessa in entrambi i casi. E l'uno è tanto (o poco) costitutivo del sé quanto l'altro, poiché in entrambi i casi è richiesto un sé per decidere che cosa fare.

L'argomento della Korsgaard presuppone (1) che affinché il sé prenda delle decisioni esso deve prenderle in accordo con un principio universale; e quella stessa presupposizione presuppone a sua volta (2) che l'agire sulla base di un principio sia in qualche modo costitutivo del sé. Io sto rifiutando entrambe queste affermazioni. Kant aveva torto: l'azione libera non richiede l'agire in accordo con una legge auto-creata. E il sé che si impegna nell'azione libera non richiede principi universali per essere un sé. Al contrario, sia il comportamento coerente che quello capriccioso all'interno del divario, come ho sostenuto nel capitolo 3, richiedono un sé *preesistente*. In breve, affinché i miei atti siano liberi e liberamente scelti da me non è necessario alcun requisito logico che esemplifichi dei principi universali. I miei atti possono essere assolutamente capricciosi ed essere ancora degli atti liberi.

Non è questa la sede per provare a formulare una diagnosi completa del potente argomento filosofico della Korsgaard, ma – sebbene troppo brevemente – penso che l'origine del suo errore sia dato dal fatto che lei vuole colmare il divario. Lei vuole che il sé sia la causa delle azioni libere. Se si accetta questo requisito allora, sulla base di certe assunzioni naturali, seguirà tutto il resto. I passaggi sono i seguenti: (1) le azioni libere sono causate dal sé. (2) Ma il sé, nel causarle, deve riferirsi a una legge e le sole leggi alle quali può riferirsi sono auto-create. (3) Nel creare una legge il sé crea sé stesso come sé.

Io respingo tutte queste considerazioni. Se con “causa” intendiamo le “condizioni causalmente sufficienti” allora le azioni libere non sono causate da nulla. Questo è ciò che le rende libere. Per dirla in maniere più precisa: ciò che rende un'azione libera a livello psicologico è il fatto che essa non deve avere antecedentemente condizioni causali psicologiche sufficienti (per l'argomentazione si veda il capitolo 3). Il sé *compie* l'atto ma non *causa* l'atto. Niente colma il divario.

IV. L'universalità del linguaggio e l'altruismo forte

Facciamo ora il punto della situazione. Stavamo cercando di rispondere alla domanda: dato che la Creatura è stata programmata per preoccuparsi dei propri interessi, c'è qualche requisito logico in questo fatto che fa sì che essa debba essere attenta agli interessi e ai bisogni delle altre persone? Le parole "altruista" ed "egoista" vengono fatte circolare senza una definizione molto chiara, proviamo quindi a definirle per la nostra discussione. In un certo senso un egoista è qualcuno che si cura solamente dei propri interessi e un altruista è qualcuno che si prende cura degli interessi degli altri. Ma questa definizione oscura una distinzione cruciale. Un altruista potrebbe essere qualcuno che è naturalmente incline a prendersi cura degli interessi degli altri, ma per un altruista di questo tipo l'agire in maniera altruista è semplicemente agire sulla base di una inclinazione tra le altre. A questa persona piace aiutare gli altri così come le piace bere, per esempio. Chiamiamo questo il senso debole dell'"altruismo". Ma c'è un senso più forte di "altruismo" che stiamo cercando di cogliere. Un altruista in questo senso è qualcuno che riconosce gli interessi degli altri come una ragione valida per agire anche nei casi in cui egli non ha questa inclinazione. La domanda è la seguente: vi sono ragioni per l'azione razionalmente vincolanti e altruistiche *indipendenti dal desiderio*? Un altruista in senso forte è colui che riconosce che vi sono ragioni razionalmente vincolanti e indipendenti dal desiderio per agire nell'interesse degli altri. Sia Nagel che Kant-Korsgaard hanno fornito delle argomentazioni a sostegno del requisito razionale dell'altruismo in senso forte. I socio-biologi rispondono alla domanda solamente in senso debole. Io ho respinto sia l'argomentazione di Nagel che quella di Kant-Korsgaard. Ma penso che le loro conclusioni siano giuste e penso che Kant-Korsgaard abbia ragione nel notare che il problema riguarda la generalità. Dando per scontato che la Creatura e noi stessi abbiamo delle ragioni per comportarci egoisticamente, esiste un requisito di generalità che estenderebbe queste ragioni ad altre persone in modo da vincolare il nostro comportamento? Io credo che vi sia un tale requisito.

La generalità richiesta per sostenere l'altruismo forte è già insita nella struttura del linguaggio. In che modo, esattamente? Procediamo per vari passaggi per vedere come il linguaggio introduce forme di generalità richieste razionalmente. Sia io che il mio cane possiamo vedere che c'è un uomo alla porta, cioè abbiamo entrambi l'esperienza visiva che a parole io descrivo come il "vedere che c'è un uomo alla porta". Ma c'è una grande

differenza in questo se io *dico* di vedere un uomo alla porta in un linguaggio che mi impegna in un imperativo categorico semantico che non ha alcun analogo nel cane. Quando dico “quello è un uomo”, mi impegno in una affermazione secondo cui ogni entità uguale a quella, in senso rilevante, può anche essere correttamente descritta come “un uomo”. Per usare la terminologia kantiana: gli asserti sono vincolati dall’imperativo categorico semantico: e affermare così che la massima della vostra asserzione può essere intesa come una legge universale alla quale tutti i parlanti sono vincolati. E la massima è fornita dalle condizioni di verità della proposizione asserita. In questo caso: un oggetto che abbia queste particolari caratteristiche soddisfa le condizioni di verità per un “uomo”.

Quando si esprime un asserto della forma a è un F , la razionalità richiede che si sia capaci di volere che chiunque in una simile situazione dovrebbe asserire che a è un F . Cioè, poiché il predicato è generale, la sua applicazione richiede che qualsiasi persona che lo usi ne riconosca la generalità. Qualsiasi persona che usa il linguaggio, secondo la formulazione kantiana, deve essere capace di volere una legge universale della sua applicazione a casi simili.⁶

Inoltre, questo imperativo, diversamente da altri imperativi kantiani, soddisfa in effetti la condizione di Kant secondo la quale il bugiardo o il disonesto quando tenta di volere che la sua massima sia una legge universale cadrebbe in qualche tipo di auto-contraddizione. Supponiamo che io stia mentendo quando dico “quello è un uomo”; in questo caso non posso volere una legge universale per la quale chiunque in una situazione simile direbbe “quello è un uomo”, poiché se lo facesse, la parola “uomo” cesserebbe di avere il significato che ha. Questo significa che io non posso collegare coerentemente la mia volontà per la quale il mio enunciato è una bugia con la mia volontà che il contenuto semantico si applichi universalmente in accordo con l’imperativo categorico semantico.

Per esporre questo punto senza l’apparato kantiano possiamo dire che qualsiasi asserto di un parlante S della forma a è un F impegna S a una generalizzazione universale: per ogni x , se x è di un tipo identico in senso rilevante ad a , allora x è correttamente descritto come “ F ”. Non stiamo qui parlando delle relazioni di implicazione tra

⁶ Certamente né nel mio caso né in quello di Kant la facoltà di volere una legge universale richiede che l’agente pensi che sarebbe una buona cosa che tutti si comportassero allo stesso modo. Non è questo il punto. Alla fin fine sarebbe molto noioso e stancante se chiunque nella mia situazione fosse portato a dire “quello è un uomo”. Il punto che riguarda l’imperativo categorico è logico; non c’è alcuna assurdità logica nel mio desiderare la massima dell’azione come una legge universale che lega insieme tutti i parlanti.

proposizioni, ma piuttosto di ciò su cui un parlante si impegna quando effettua degli atti linguistici.

Inoltre, il requisito di generalità si applica alle altre persone. Se sono impegnato a riconoscere istanze simili, come altri casi di uomini, il mio impegno in un linguaggio pubblico mi richiede di considerare che anche altre persone debbano riconoscere questo, e casi simili, come casi di uomini. La generalità è insita nella struttura del linguaggio stesso e infatti quando la si applica al linguaggio comunque la mettiamo è come se passassimo da ciò che è ciò che dovrebbe essere. Dal fatto che un oggetto viene veramente descritto come “un uomo”, segue che si dovrebbero accettare anche oggetti simili come uomini, e che altre persone dovrebbero sia accettare questo come un uomo che altri oggetti simili come uomini. È impossibile usare il linguaggio senza questi impegni. Ho esposto tutto questo con una terminologia altisonante, ma si tratta di una conseguenza banale della natura del linguaggio e degli atti linguistici.

Il modo in cui passiamo dalla generalità alle ragioni per l'azione nella forma di un altruismo forte è semplicemente notando che il requisito della generalità che vale per predicati come “uomo”, “cane” e “montagna” vale anche per “egli ha una ragione per l'azione” e altri motivatori del genere. Mostrerò tutto questo con un esempio. Supponiamo che io abbia un dolore e che cerchi di alleviare questo dolore. C'è una differenza tra il mio cercare di alleviare il dolore e il mio cane che allevia il suo dolore leccandosi la ferita. Qual è la differenza? Ebbene, se non altro questa: io posso ricondurre il mio dolore a certe generalizzazioni universali, semplicemente caratterizzandolo con una parola come “dolore”. Cioè la stessa caratteristica che abbiamo trovato nella discussione della parola “uomo” verrà applicata anche alla parola “dolore”. Se io affermo “questo è un dolore” sono impegnato alla condizione “per tutti gli x , se x è simile a questo in senso rilevante, x è un dolore”.

La generalità del linguaggio date certe assunzioni di senso comune relative ai miei auto-interessi, genererà un altruismo forte. Per prima cosa porrò la questione in forma intuitiva per poi reconsiderarla in forma semantica. Intuitivamente sembra ragionevole supporre che se io ho un dolore ho una ragione per volerlo alleviare. Il mio sentire questo grado di dolore implica il sentire la necessità di alleviarlo. Il mio bisogno di alleviare il dolore è per me una ragione per alleviare il mio dolore e credo che anche gli altri, se ne hanno la capacità e l'opportunità, abbiano una ragione per aiutarmi ad alleviare il dolore. Ma io non posso credere che essi abbiano una ragione per aiutarmi, senza impegnare me

stesso a credere che nella stessa situazione, ma con i pronomi invertiti, sarei portato a riconoscere di avere una ragione per aiutare loro. È razionale da parte mia desiderare che essi mi aiutino, per la ragione che io ora ho bisogno di aiuto. Ma allora, coerentemente, quando essi hanno bisogno di aiuto io sono impegnato a riconoscere l'esistenza del loro bisogno come una ragione per aiutarli.

Il modo in cui la generalità del linguaggio produce un altruismo forte è il seguente:

1. Io ho un dolore, quindi dico "io ho un dolore". Quando affermo "io ho un dolore" sono impegnato, per il requisito della generalità, a riconoscere che in una situazione simile voi avreste un dolore. Poiché "dolore" è un termine generale del linguaggio, le condizioni di verità si applicano indifferentemente sia a voi che a me. Io sono impegnato ad applicare l'espressione aperta "X ha un dolore" a qualsiasi oggetto che soddisfi esattamente queste condizioni.
2. Il mio dolore crea un bisogno. Poiché ho un dolore io ho bisogno di aiuto. Io sono consapevole sia del mio dolore che del mio bisogno. Così io affermo "ho bisogno di aiuto perché ho un dolore". Ora notate che questo non deve essere interpretato come una richiesta di aiuto. Non è un atto linguistico indiretto; piuttosto è un enunciato proferito da me riguardo a me stesso. Lo stesso requisito di generalità si applica ancora. Io ora sono impegnato a riconoscere che in una situazione simile, con i pronomi invertiti, se voi avete un dolore avreste bisogno di aiuto. Sono impegnato ad applicare l'espressione aperta "X ha bisogno di aiuto perché X ha un dolore" in ogni situazione di identico tipo.
3. Io ho un dolore ed ho bisogno di aiuto e credo che il mio bisogno di aiuto sia una ragione per voi per aiutarmi. Così supponiamo che io dica "poiché ho un dolore ed ho bisogno di aiuto, voi avete una ragione per aiutarmi". E' in vigore lo stesso requisito di generalità. Io sono impegnato all'universale per qualsiasi situazione che sia di tipo identico a questa, in senso rilevante:

per tutti gli x e per tutti gli y , se x ha un dolore e x ha bisogno di aiuto poiché x ha un dolore, y ha una ragione per aiutare x .

Ma questo mi impegna a riconoscere che quando voi avete un dolore io ho una ragione per aiutarvi. Notate che qui stiamo parlando degli impegni dei parlanti nell'effettuare degli atti linguistici. Non stiamo, in questo punto, considerando la verità o le relazioni di implicazione tra le proposizioni; piuttosto ci stiamo preoccupando di ciò verso cui il parlante è impegnato quando fa un'asserzione di questa forma.

Il punto essenziale della presente discussione è che una volta che abbiamo programmato la Creatura nel modo che ho descritto, ovvero, che in aggiunta alle capacità mentali di base le abbiamo dato il divario, l'auto-interesse e il linguaggio, allora le abbiamo già fornito una base logica sufficiente per un altruismo forte. Notate, inoltre, che non richiediamo una metafisica del dovere. Non è necessario alcun mondo noumenico o un imperativo categorico kantiano. Tutto ciò che questo argomento richiede è che noi, altre persone e la Creatura possiamo parlare in italiano o in altri linguaggi e che esprimiamo asserzioni ragionevoli di auto-interesse. Noi affermiamo, per esempio, che i nostri bisogni possano essere una ragione per qualcun altro per aiutarci.

Ma perché non potrei oppormi a questo argomento dicendo, per esempio, che il mio è un caso speciale? Io merito un trattamento speciale non concesso ad altri. Qualcuno potrebbe sempre sostenere una pretesa del genere, ma fare questo va al di là della semantica degli indessicali. Non c'è niente nella semantica di "io", "tu", "egli", ecc. che si opponga alla generalità delle condizioni di verità di "dolore", "bisogno", "ragione", ecc. Non sto qui cercando di eliminare la possibilità di trattamenti speciali o la cattiva fede. La storia del mondo è piena di persone, tribù, classi, nazioni ecc. che ingannano pretendendo un diritto a privilegi speciali e niente di ciò che dico potrà fermare queste persone dall'ingannare. La mia tesi è piuttosto che il vincolo di universalità che ci porta dall'egoismo all'altruismo forte è già insito nell'universalità del linguaggio. Tutto ciò che dobbiamo assumere è che la Creatura abbia certe attitudini ragionevoli di auto-interesse che riguardano le sue relazioni con gli altri esseri coscienti e che sia preparata ad esprimerle in un linguaggio. Una volta che la Creatura o chiunque altro è preparata a dire "voi avete una ragione per aiutarmi perché io ho un dolore ed ho bisogno di aiuto" allora essa è impegnata, in situazioni di tipo identico, ad applicare i quantificatori universali all'espressione aperta "y ha una ragione per aiutare x

perché *x* ha un dolore e ha bisogno di aiuto”, poiché l’uso dei termini generali impegna il parlante all’applicazione di quei termini alle situazioni che condividono le caratteristiche generali della situazione iniziale. Il linguaggio è per sua stessa natura generale.

In riferimento a chi contesta questa conclusione, ritengo che tale contestazione provenga da un altro errore diffuso della nostra cultura: l’idea che il linguaggio non possa avere tutta questa grande importanza, dal momento che è costituito da semplici parole. Come può la semplice pronuncia di parole impegnarmi verso qualcosa? Ho incontrato questa stessa resistenza molto tempo fa quando dimostrai in che modo derivare il “dovrebbe” dall’“è”⁷. Molti commentatori osservarono che il semplice fatto di pronunciare delle parole non poteva impegnarmi verso nulla. Ci deve essere un qualche principio morale ulteriore o una qualche approvazione delle istituzioni del linguaggio. O qualcosa del genere!

Dirò di più su questi argomenti nel prossimo capitolo ma, per ora, possiamo dire che il problema non sta nel vedere come il pronunciare delle parole possa impegnarmi ma, piuttosto, nel vedere come qualcosa *che non sia* il pronunciare delle parole possa impegnarmi. Le forme paradigmatiche dell’impegno verso dei corsi d’azione si trovano nell’esecuzione di atti linguistici.

V. Conclusioni

In questo capitolo mi sono posto tre obiettivi principali. Ho provato a descrivere alcune caratteristiche speciali delle ragioni per l’azione; ho provato a descrivere quali caratteristiche sono necessarie perché un sé-agente sia capace di razionalità; ho provato a derivare i principi di un altruismo forte dall’universalità del linguaggio, assieme alle assunzioni di senso comune che riguardano l’auto-interesse.

Quali implicazioni hanno questi argomenti e quelli dei capitoli precedenti per il modello classico della razionalità? Il modello classico, potremmo dire, è progettato per degli scimpanzé estremamente intelligenti. Non ha a che fare con certe caratteristiche speciali della razionalità umana, specialmente quelle caratteristiche speciali che sono rese possibili e sono in effetti richieste dall’istituzione del linguaggio. Finora ho discusso tre modi in cui il

⁷ John R. Searle, “How to Derive ‘Ought’ from ‘is’”, *Philosophical Review*, 73, gennaio 1964, pp. 43-58.

modello classico fallisce nello spiegare certe caratteristiche diffuse del prendere decisioni razionali.

1. Il modello classico non può spiegare il ragionamento prudentiale a lungo termine, dove le considerazioni prudentiali non sono rappresentate nell'insieme motivazionale corrente del sé in questione. L'esempio della studentessa fumatrice in Danimarca è stato utilizzato per illustrare questo punto.
2. Il modello classico non può spiegare la razionalità riconoscitiva in cui il sé cosciente riconosce un motivatore indipendente dal desiderio in quanto capace di fornire una ragione per l'azione. Lo scimpanzé può, presumibilmente, riconoscere fonti immediate di pericolo o oggetti desiderabili come il cibo, ma lo scimpanzé non può riconoscere allo stesso modo entità fattive come gli obblighi, gli impegni e i bisogni a lungo termine.
3. Il modello classico non può spiegare le implicazioni dell'universalità del linguaggio. Data questa universalità, insieme a certe assunzioni naturali che riguardano i tipi di ragione che ognuno accetta per sé, segue come conseguenza l'altruismo forte.

Nel prossimo capitolo ci occuperemo di:

4. La creazione intenzionale di ragioni indipendenti dal desiderio da parte delle azioni intenzionali coscienti del sé.

P 68

6.

IN CHE MODO CREIAMO DELLE RAGIONI PER L'AZIONE INDIPENDENTI DAI DESIDERI

I. LA STRUTTURA DI BASE DEL DIVARIO (GAP)

La capacità più notevole della razionalità umana e il modo in cui essa si differenzia dalla razionalità delle scimmie, è la capacità di creare e di agire sulla base di ragioni per l'azione indipendenti dai desideri. La creazione di tali ragioni è sempre una questione che riguarda un agente *che impegna* se stesso in vari modi. Il modello classico non può spiegare né l'esistenza né la forza vincolante razionale di queste ragioni e, infatti, gran parte degli autori appartenenti alla tradizione del modello classico negano l'esistenza di tali cose. Abbiamo visto che la prudenza a lungo termine costituisce già una difficoltà per il modello classico, poiché sulla base di questo modello un agente può agire razionalmente solamente sulla base del desiderio che ha proprio in quel momento. Abbiamo visto, nel caso della fumatrice in Danimarca, che può essere un requisito della razionalità il fatto che un agente che manchi del desiderio proprio nel momento di agire sulla base di considerazioni prudenziali a lungo termine, abbia nondimeno una ragione per fare ciò che fa. Il modello classico non riesce a spiegare questo fatto. Sulla base del modello classico, il soldato che si butta su una granata innescata per salvare le vite dei suoi compagni è esattamente nella stessa situazione, da un punto di vista razionale, del bambino che sceglie il cioccolato anziché la vaniglia quando decide il gusto del gelato. Il soldato preferisce la morte, il bambino preferisce il cioccolato. In ogni caso, la razionalità è una questione che riguarda l'aumento della probabilità di giungere ad un gradino più alto nella scala delle preferenze.

Non vorrei tuttavia che casi eroici come questo facessero sembrare che la creazione di e l'agire sulla base di ragioni per l'azione indipendenti dal desiderio sia qualcosa di bizzarro o inusuale. A me sembra che noi creiamo delle ragioni indipendenti dal desiderio

più o meno ogniqualvolta apriamo la bocca per parlare. In questo capitolo andremo ad esaminare un'ampia classe di casi in cui noi creiamo queste ragioni. È importante dire all'inizio esattamente qual è il problema. In un senso molto ampio di "volere" e "desiderare", ogni azione intenzionale è un'espressione o una manifestazione di un volere o di un desiderio di compiere quell'azione. Certamente, quando vado dal dentista per farmi trapanare un dente non ho il desiderio, la gran voglia, la passione, la bramosia, la *Sehnsucht*, il piacere o l'inclinazione di farmelo trapanare; ma allo stesso tempo, proprio in quel momento, quello è ciò che voglio fare. Voglio farmi trapanare il dente. Un desiderio del genere è un desiderio motivato o secondario. È motivato dal mio desiderio di farmi curare il dente. Ora, poiché ogni azione intenzionale è l'espressione di un desiderio, sorge la domanda: da dove provengono questi desideri? Secondo il modello classico ci possono essere soltanto due possibilità: o l'azione è un'azione che desidero compiere di per sé o essa è un'azione che compio in funzione di qualche altro desiderio che possiedo. Sto bevendo questa birra perché voglio bere questa birra oppure la sto bevendo per soddisfare qualche altro desiderio; ad esempio, credo che questo sarebbe di beneficio per la mia salute ed io desidero migliorarla. Non ci sono altre possibilità. Secondo questa spiegazione la razionalità è una questione che riguarda interamente la soddisfazione dei desideri.

Suona un po' grossolano affermare che ogni azione razionale viene compiuta per soddisfare un desiderio ed è quindi interessante osservare come i teorici della tradizione classica procedano con grande pesantezza quando arriva il momento di descrivere la motivazione. In che modo, esattamente, essi descrivono la motivazione razionale? Bernard Williams, il quale pensa che non vi siano ragioni esterne e che ogni atto razionale debba rifarsi a qualcosa che sta nell'insieme motivazionale *S* dell'agente, sostiene quanto segue riguardo ai contenuti di *S*:

Io ho parlato di *S* fondamentalmente in termini di desideri e, da un punto di vista formale, tale espressione può essere usata per tutti gli elementi di *S*. Ma c'è il rischio che questa terminologia ci faccia dimenticare che *S* può anche contenere, per esempio, disposizioni valutative, modelli di reazione emotiva, legami personali, nonché progetti di vario tipo – come potremmo astrattamente chiamarli – in cui si concretizza *l'impegno dell'agente* (il corsivo è mio)¹.

¹ Bernard Williams, "Internal and External Reasons" in *Moral Luck*, Cambridge, Cambridge University Press, 1981, p. 105, tr. it. "Ragioni interne ed esterne" in *Sorte morale*, Il Saggiatore, Milano, 1987, p. 138.

Una biforcazione simile si trova nella caratterizzazione di Davidson delle “pro-attitudini”. Ecco che cosa dice: “quando qualcuno fa qualcosa per una ragione, lo può caratterizzare dicendo che (a) ha qualche tipo di pro-attitudine (o atteggiamento favorevole) verso azioni di un certo tipo, e (b) crede (o sa, percepisce, si accorge, ricorda) che la sua azione sia di quel tipo”². E del suo insieme di pro-attitudini egli elenca le seguenti. Era qualcosa che l’agente “voleva, desiderava, apprezzava, aveva caro, riteneva *doveroso*, benefico, *obbligatorio* o condivisibile” (ibid., il corsivo è mio). Il problema con questo elenco, così come con quello di Williams, è che esso offusca la distinzione tra ragioni per l’azione dipendenti dai desideri e indipendenti dai desideri. Offusca la distinzione tra cose che si *vogliono* fare e cose che si *devono* fare indipendentemente dal fatto che le si voglia fare o meno. Volere o desiderare qualcosa è una cosa sola, da considerare in maniera diversa rispetto a qualcosa di “obbligatorio” o ad un “impegno” (*commitment*) che si devono fare indipendentemente dal desiderio. Perché Williams e Davidson non ci dicono che cosa sono un impegno o un obbligo? Sono semplicemente un altro desiderio, “formalmente” parlando?

Io penso che la ragione per cui entrambi gli autori sembrano essere in lotta in questo caso è che essi vogliono assimilare le ragioni per l’azione indipendenti dal desiderio, che ovviamente esistono, ai desideri. E fanno questo affermando che se costruiamo un insieme che comprenda i desideri in senso in generale, allora gli impegni, gli obblighi di una persona ecc. saranno davvero membri dello stesso insieme come desideri. Io penso che questo offuschi la distinzione cruciale che sto cercando di porre tra desideri e ragioni per l’azione indipendenti dal desiderio. Perché ci sarebbe una tale distinzione? Sicuramente, la gente può *volere* adempiere ai propri obblighi e mantenere le promesse. Sì, ma questa non è la stessa cosa del volere un gelato al cioccolato. Io voglio del cioccolato e voglio mantenere la mia promessa. Qual è la differenza? *Nel caso della promessa il desiderio è derivato dal riconoscimento della ragione indipendente dal desiderio, cioè l’obbligo. La ragione è precedente al desiderio e sta alla base del desiderio. Nel caso del cioccolato il desiderio è la ragione.*

I punti in discussione in questo capitolo sono l’esistenza, la natura, la creazione e il funzionamento delle ragioni per l’azione indipendenti dal desiderio, che soddisfano le seguenti condizioni di adeguatezza:

² Donald Davidson, “Actions, Reasons, and Causes” ristampato in A. White (ed.), *The Philosophy of Action*, Oxford, Oxford University Press, 1968, p. 79, tr.it. “Azioni, ragioni e cause”, in *Azioni ed eventi*, il Mulino, Bologna 1992, pp. 42.

1. La spiegazione deve essere completamente naturalistica. Deve cioè mostrare in che modo la creazione e il funzionamento di queste ragioni sono possibili per animali biologici quali noi siamo. Noi siamo differenti dagli scimpanzé, ma le nostre capacità sono un'estensione naturale delle capacità di qualsiasi altro primate. Non deve esserci alcun appello ad alcunché di trascendentale, non biologico, noumenico o soprannaturale. Stiamo semplicemente parlando di certe capacità di animali biologici come noi stessi.
2. Deve specificare l'apparato che ci consente di creare delle ragioni per l'azione indipendenti dal desiderio.
3. Deve spiegare in che modo, all'interno di quell'apparato, le persone creano tali ragioni. Ho bisogno di stabilire esattamente la struttura logica dell'intenzionalità che sottostà alla creazione delle ragioni per l'azione indipendenti dal desiderio.
4. Deve spiegare in che modo la razionalità da sola rende queste ragioni vincolanti per l'agente. Per quale ragione razionale l'agente dovrebbe prendere in considerazione i suoi impegni e i suoi obblighi? Perché non può semplicemente ignorarli?
5. Deve spiegare in che modo il riconoscimento razionale di tali ragioni è sufficiente per costituire la motivazione: in che modo entità di questo tipo possono fondare razionalmente i desideri secondari se esse stesse sono indipendenti dal desiderio.
6. Deve spiegare in che modo l'apparato e l'intenzionalità usata per rispondere alle condizioni (1) – (5) è sufficiente sia per la creazione che per l'operazione di tali ragioni. Non c'è bisogno di alcun aiuto da parte di principi generali, regole morali ecc. Cioè, la risposta alle condizioni (1) – (5) deve spiegare in che modo vengono create le ragioni per l'azione indipendenti dal desiderio e in che modo esse funzionino senza l'assistenza di principi morali essenziali. Le ragioni indipendenti dal desiderio devono essere, per così dire, auto-sufficienti.

Chiunque abbia familiarità con la storia della filosofia occidentale penserà che ho assegnato a me stesso un compito scoraggiante. Molti critici descrivono questo tipo di impresa come se si trattasse di tirar fuori un coniglio dal cappello. Ma io credo che in realtà, se riusciamo a dimenticarci del modello classico e dell'intera tradizione che esso comprende, la risposta ai nostri problemi, sebbene complessa nel dettaglio, sia piuttosto semplice nella sua struttura di base.

È importante, comunque, fornire una spiegazione al livello corretto, poiché ci sono differenti livelli ai quali queste domande possono dare risposta. C'è il livello "fenomenologico" nel quale descriviamo come le cose appaiono all'agente quando egli è coinvolto in un comportamento razionale socialmente impegnato, e c'è il livello sociale o "societario" nel quale discutiamo le istituzioni sociali utilizzate nella creazione di tali ragioni per l'azione indipendenti dal desiderio, quando spieghiamo in che modo tali istituzioni sono strutturate e quali funzioni esse ricoprono nell'intera società.

Dirò qualcosa su questi livelli più avanti, ma ora vorrei iniziare con la discussione del livello di intenzionalità più semplice e di base. Questo è, per così dire, il livello atomico che è precedente ai livelli molecolari della fenomenologia e della sociologia. Negli ultimi paragrafi parlerò più in dettaglio dell'impegno, della sincerità e della falsità e del ruolo specifico delle istituzioni umane. Ma all'inizio è importante fare chiarezza sulle forme più semplici e primitive degli impegni umani. Quali sono le condizioni di soddisfazione dei fenomeni intenzionali implicati nella creazione di impegni? Supponiamo di avere un parlante e un ascoltatore che siano entrambi in grado di parlare e di comprendere un linguaggio comune. Supponiamo che essi siano i responsabili della capacità di formulare degli enunciati, delle richieste, delle promesse ecc. Nei tipi più semplici di atti linguistici, in cui il parlante esprime ad esempio un'asserzione, una richiesta o una promessa, egli impone delle condizioni di soddisfazione su altre condizioni di soddisfazione. In che modo, esattamente? Consideriamo l'esempio dell'esprimere un'asserto con particolare attenzione e osserviamo che cosa accade. Supponiamo che il parlante pronunci una frase, per esempio "sta piovendo", e supponiamo che egli intenda esprimere l'asserto che sta piovendo. La sua intenzione-in-azione è, in parte, produrre l'espressione "sta piovendo". Tale espressione è una delle condizioni di soddisfazione della sua intenzione. Ma se egli non sta semplicemente esprimendo l'enunciato, ma sta effettivamente *dicendo che* sta piovendo, se effettivamente *vuole significare* che sta piovendo, allora deve intendere che l'espressione soddisfi delle condizioni di verità, le condizioni di soddisfazione con direzione di adattamento verso il basso che sta piovendo. Cioè, il significato della sua intenzione è di imporre delle condizioni di soddisfazione (per esempio, le condizioni di verità) su altre condizioni di soddisfazione (l'enunciato). Il suo enunciato, ora, ha una funzione di *status*, rappresenta, in maniera vera o falsa, lo stato del tempo atmosferico. Ed egli non è neutrale di fronte alla verità o alla falsità, perché la sua affermazione è una affermazione di verità. *Tale*

imposizione di questo tipo di funzione di status, di condizioni di soddisfazione su condizioni di soddisfazione, è già un impegno. Perché? Perché l'asserzione era un'azione libera, intenzionale, del parlante. Egli si è assunto la responsabilità di sostenere che sta piovendo e si impegna così alla verità della proposizione asserita. Quando egli, intenzionalmente, impone delle condizioni di soddisfazione su altre condizioni di soddisfazione, nella forma di un'asserzione, egli si assume la responsabilità che queste condizioni vengano soddisfatte. E questo impegno è già una ragione per l'azione indipendente dal desiderio. Il parlante ha creato in questo caso una ragione per accettare le conseguenze logiche della sua asserzione, per non negare ciò che ha detto, per essere in grado di fornire evidenza o giustificazione di quanto ha detto e per parlare in modo sincero. Tutto questo è il risultato delle regole costitutive dell'espressione degli enunciati e il parlante, quando impone condizioni di soddisfazione su condizioni di soddisfazione, invoca tali regole. La creazione degli impegni crea delle ragioni per l'azione indipendenti dal desiderio, e l'impegno è già insito nella struttura dell'atto linguistico. Nel formulare un'asserzione il parlante presenta una proposizione con una direzione di adattamento verso il basso. Ma così facendo, crea un impegno che ha direzione di adattamento verso l'alto. La sua asserzione che sta piovendo sarà vera o falsa a seconda del fatto che stia davvero piovendo. Ma l'impegno che egli ha creato verrà soddisfatto solo se il mondo è come egli sostiene che sia, solo se sta piovendo.

Finora abbiamo considerato soltanto le asserzioni ma, in realtà, tutte le forme tipiche di atti linguistici con contenuti proposizionali implicano la creazione di ragioni per l'azione indipendenti dal desiderio, poiché l'imposizione intenzionale di condizioni di soddisfazione impegna o obbliga il parlante in vari modi. Persino le richieste e gli ordini, sebbene il loro contenuto proposizionale si riferisca a condizioni imposte all'ascoltatore piuttosto che al parlante, impegnano il parlante in vari modi. Se io vi ordino di lasciare la stanza mi impegno a permettervi di lasciare la stanza e a volere che voi lasciate la stanza, per esempio.

Che cos'è, allora, un impegno? Il modo per rispondere a questa domanda è analizzare la struttura logica degli impegni. Gli impegni sono entità fattive che soddisfano la nostra condizione per le ragioni per l'azione. Un impegno ha un contenuto proposizionale e una direzione di adattamento verso l'alto. Così, se ho l'impegno di andare a San Jose la prossima settimana, il contenuto proposizionale è "che io vada a San Jose la prossima settimana" e la direzione di adattamento è verso l'alto. L'impegno è soddisfatto solo se il mondo cambia fino ad accordarsi con il contenuto dell'impegno, solo se effettivamente vado a San Jose.

Senza tentare di fornire delle “condizioni necessarie e sufficienti” si può dire questo: un impegno è l’adozione di un corso di azione o di una certa condotta (o di altri contenuti intenzionali; si può, per esempio, essere vincolati a credenze e desideri) dove la natura dell’adozione fornisce una ragione per effettuare il corso d’azione. Così, per esempio, io sono impegnato a fare filosofia. E questo impegno mi fornisce una ragione per farlo anche nei momenti difficili, quando le cose non vanno bene. Allo stesso modo si può essere impegnati nei confronti della Fede Cattolica o nei confronti del Partito Democratico. Quando Sally dice che Jimmy non vuole “impegnarsi” intende dire che lui non vuole adottare una condotta che gli fornirà una ragione per continuare in certi comportamenti e attitudini. Ragioni di questo tipo sono indipendenti dal desiderio, sebbene questo venga oscurato dal fatto che i tipi di impegni che ho descritto sono impegni a fare delle cose che si vorrebbero comunque fare. In questo capitolo ci occuperemo principalmente di una forma speciale di impegno, in cui si crea un impegno verso un’altra persona attraverso l’imposizione di condizioni di soddisfazione su condizioni di soddisfazione.

Dopo avere analizzato la struttura logica degli impegni, risulta più facile vedere in che modo possiamo creare un impegno quando compiamo un atto linguistico. Per esempio, si può impegnare se stessi ad una certa condotta semplicemente adottando una ferma intenzione a proseguire in quella condotta, ma in questo momento vorrei considerare la classe di impegni che vengono creati pubblicamente, che sono normalmente diretti ad altre persone. Possiamo creare un impegno del genere per noi stessi imponendo delle condizioni di soddisfazione su qualche altra entità. È più difficile vedere come questo funzioni per gli assertivi che non per i commissivi, perché nel caso di un’asserzione noi imponiamo all’enunciato delle condizioni di soddisfazione con direzione di adattamento verso il basso, cioè facciamo un’asserzione di verità. Ma nel fare un’asserzione di verità imponiamo degli obblighi anche su *noi stessi*. Nel fare un’asserzione ci prendiamo la *responsabilità* della verità, della sincerità e dell’evidenza. E tali responsabilità, come gli impegni in generale, hanno una direzione di adattamento verso l’alto. Queste responsabilità sono soddisfatte solo se il mondo è tale che l’enunciato è vero, se il parlante è sincero e se fornisce un’evidenza per l’asserzione.

Ma perché questi impegni, obblighi e responsabilità vincolano l’agente? Perché egli, razionalmente, non li può semplicemente ignorare? Perché non sono costrutti sociali come altri? Perché il parlante si trova in una relazione speciale con le sue asserzioni, nel senso che

egli le ha create come propri impegni. Assumendo tali impegni egli ha vincolato se stesso liberamente e intenzionalmente. Egli può essere indifferente nei confronti della verità dell'asserzione di qualcun altro, perché non ha impegnato se stesso. Egli non può essere indifferente nei confronti della verità delle proprie asserzioni, precisamente perché questi sono i suoi impegni.

Ma come può un tale impegno astratto, indipendente dal desiderio, dare origine ad un desiderio secondario? Come può mai motivarlo? Ebbene, chiedetevi: in che modo l'evidenza, la prova e perfino la verità stessa fanno sì che qualcuno creda qualcosa alla quale non vuole credere? Per esempio, molte persone non vogliono credere al teorema di Gödel perché esso ha distrutto il loro progetto di ricerca. Ma qualora essi riconoscessero la validità della dimostrazione, razionalmente parlando, non avrebbero scelta. Riconoscere la validità della dimostrazione è già riconoscere una ragione per accettarla e riconoscere una ragione per accettarla è già riconoscere una ragione per volerla accettare. La lezione contenuta in questo caso, e in altri che considereremo, è che le ragioni indipendenti dal desiderio forniscono delle motivazioni come qualsiasi altra ragione. Una volta che si riconosca qualcosa come una ragione valida per agire, ovvero, una volta che si riconosca un'entità fattiva in cui voi siete il soggetto e con una direzione di adattamento verso l'alto, l'avrete già riconosciuta come un motivo per voler fare la cosa che vi siete impegnati a fare. Il mio desiderio di dire la verità o di mantenere la mia promessa è derivato dal fatto che io riconosco che esprimo un enunciato o che ho fatto una promessa, che enunciati e promesse creano impegni e obblighi, e che mi viene richiesto di adempiere ai miei impegni e ai miei obblighi, nello stesso modo in cui il mio desiderio di farmi trapanare il dente è derivato dal mio riconoscimento che esso ha bisogno di essere curato e dal mio desiderio di occuparmi della mia salute.

Le persone tendono ad assumere che il modo in cui le ragioni dipendenti dal desiderio motivano i desideri secondari non sia problematico. Ma il modo in cui le ragioni dipendenti dal desiderio motivano i desideri secondari è altrettanto problematico del modo in cui le ragioni indipendenti dal desiderio motivano tali desideri. Io riconosco che il mio desiderio di farmi curare il dente è una ragione per farmelo trapanare, e perciò una ragione per volerlo fare trapanare. Riconosco anche che dovervi del denaro è una ragione per restituirvelo, e perciò una ragione per volervelo restituire. In ogni caso il riconoscimento di una valida entità fattiva in cui io sono il soggetto e con una direzione di adattamento verso

l'alto è una ragione per compiere un'azione, e perciò una ragione per voler compiere l'azione.

La difficoltà nel vedere che non c'è nulla di particolarmente problematico nel modo in cui le ragioni indipendenti dal desiderio possono fornire delle motivazioni deriva, in parte, da una tendenza presente nella nostra tradizione a pensare che la motivazione abbia a che fare con condizioni causalmente sufficienti. È un punto di debolezza della nostra tradizione supporre che qualsiasi spiegazione di una motivazione debba dimostrare in che modo l'azione è necessaria, in che modo l'agente *deve* compiere l'azione se davvero possiede le giuste ragioni per farlo. Questo errore deriva dall'incapacità di riconoscere il divario. Potrei riconoscere il mio bisogno di farmi trapanare il dente, così come potrei riconoscere il mio obbligo e tuttavia non agire ancora sulla base di una ragione. Così, in una spiegazione della forza che motiva le ragioni per l'azione indipendenti dal desiderio, non cerchiamo di dimostrare che esse causano le azioni sulla base di condizioni sufficienti. Non lo fanno. Né lo fa nessun'altra ragione razionale per l'azione.

Un passaggio essenziale per comprendere la motivazione è dato dal rendere chiare le relazioni tra il punto di vista della terza persona e il punto di vista della prima persona. Dal punto di vista della terza persona, ogni società è costituita da un insieme di strutture istituzionali e i membri di quella società sono, in modi differenti, agli occhi dei loro compagni, vincolati a strutture deontiche che stanno all'interno di quelle strutture istituzionali. Essi sono vincolati come mariti, mogli, cittadini, contribuenti ecc. Ma affermare questo non significa nulla riguardo al punto di vista della prima persona. Per quale motivo dovrei, in quanto sé cosciente, preoccuparmi di ciò che le altre persone pensano io sia vincolato o obbligato a fare? La risposta è che dal punto di vista della prima persona, io, agendo all'interno di tali strutture istituzionali, posso volontariamente e intenzionalmente creare delle ragioni per me stesso indipendenti dal desiderio. Le strutture istituzionali mi consentono di farlo, ma – e questo è il punto cruciale – gli obblighi, gli impegni e gli altri motivatori che ho creato non derivano dall'istituzione ma dal fatto che io mi impegno, intenzionalmente e volontariamente, verso tali obblighi, impegni e doveri. A causa di questo fatto io stesso, in quanto agente cosciente, posso razionalmente richiedere il riconoscimento di questi motivatori. Questo risulta ovvio nel caso delle promesse, e ugualmente vero, sebbene meno ovvio, nel caso delle asserzioni. Se esprimo la frase “io prometto”, non mi è consentito dire “sì, ho detto questo ma non vedo perché ciò costituisca il fare una

promessa”; e una volta che ho fatto la promessa non mi è consentito dire “sì, ho fatto la promessa ma non vedo perché questo mi debba impegnare”. Analogamente, se dico “sta piovendo” non mi è consentito dire “sì, ho detto questo ma non vedo perché questo costituisca l’esprimere un’asserzione”, e una volta che ho espresso un enunciato non mi è consentito dire “sì, ho espresso un enunciato, ma non vedo perché questo costituisca un impegno nei confronti della sua verità”.

Fino a questo punto ho presentato, piuttosto velocemente, una panoramica dei principali argomenti che presenterò in questo capitolo. Finora li ho discussi soltanto ad un livello di base, atomico. Più avanti arriveremo a dei livelli più elevati, e riformuleremo in maggior dettaglio l’argomento riguardante il modo in cui le ragioni indipendenti dal desiderio possono motivare le azioni. Vediamo in che modo l’esposizione degli enunciati presentata finora soddisfa le nostre condizioni di adeguatezza.

1. La spiegazione è completamente naturalistica. Le nostre abilità sono un’estensione delle abilità di animali più primitivi e, specialmente, dei primati. Le scimmie possiedono la facoltà dell’intenzionalità, ma non possiedono tale facoltà ad un secondo livello di intenzionalità in cui sé possibile imporre condizioni di soddisfazione su condizioni di soddisfazione. Esse non hanno la capacità di assumere un impegno nei confronti della verità di una proposizione secondo la quale sta piovendo imponendo condizioni di soddisfazione su condizioni di soddisfazione. Inoltre, esse non dispongono di istituzioni sociali grazie alle quali noi possiamo fare queste cose così che esse risultino riconoscibili ad altri membri della nostra specie e che, di conseguenza, ci permettono di comunicare tali impegni ad altri membri della nostra specie.

2. L’apparato che noi usiamo per la creazione di ragioni per l’azione indipendenti dal desiderio è dato dall’insieme delle regole costitutive degli atti linguistici e dalla loro realizzazione nella struttura semantica dei linguaggi umani reali. Qualsiasi linguaggio sufficientemente ricco da consentire al parlante di esprimere un’asserzione, un ordine o una promessa farà al caso nostro. Nella vita reale il parlante e l’ascoltatore sono abitualmente coinvolti in altre strutture istituzionali, come il denaro, la proprietà, le nazioni e i matrimoni.

Le strutture, sia linguistiche che non linguistiche, sono complesse. Ma non sono misteriose e le ho descritte in dettaglio altrove.³

3. Si creano ragioni per l'azione indipendenti dal desiderio imponendo condizioni di soddisfazione su condizioni di soddisfazioni. Tutte queste imposizioni sono degli impegni, e tutti questi impegni creano delle ragioni per l'azione indipendenti dal desiderio. Quando le condizioni di soddisfazione fanno riferimento al parlante, come nel caso di un giuramento o di una promessa, e il contenuto proposizionale specifica una qualche azione volontaria da parte del parlante, nell'imposizione di queste condizioni di soddisfazione c'è l'esplicita creazione di una ragione per l'azione indipendente dal desiderio. Nel caso dell'asserzione, l'impegno all'azione è soltanto implicito, ma è, nondimeno, un impegno. Imporre delle condizioni di soddisfazione all'enunciato impone degli impegni al parlante.

4. Gli impegni che voi assumete vi vincolano perché sono i *vostr*i impegni. Ovvero, dato che avete espresso l'asserzione liberamente e intenzionalmente e vi siete quindi impegnati nei confronti della sua verità, non vi è razionalmente consentito di dire che voi siete indifferenti alla verità o sincerità o coerenza, evidenza o conseguenza dell'asserto. La razionalità riconoscitiva è sufficiente. Dovete semplicemente riconoscere gli impegni che voi stessi avete creato e le loro conseguenze logiche.

5. La ragione che tali ragioni possono motivare dipende dal fatto che voi le avete create in quanto motivatori. Cioè avete creato delle entità fattive con un contenuto proposizionale che ha una direzione di adattamento verso l'alto, alla quale siete vincolati. Esercitando la volontà nell'imporre delle condizioni di soddisfazione su altre condizioni di soddisfazione, si vincola la propria volontà nel futuro direttamente a queste condizioni. Questo sarà più evidente quando considereremo le promesse, ma quasi tutti gli atti linguistici hanno un elemento di promessa. Per molto tempo i filosofi hanno provato a trattare le promesse come un tipo di asserzione. Sarebbe più esatto pensare alle asserzioni come ad una specie di promessa che riguarda un caso specifico.

6. Notate che ho fornito la risposta alle condizioni (1) – (5) senza fare alcun riferimento ad un principio esterno di tipo fondativo. Principi come “dovresti dire la verità”, “non dovresti mentire” o “dovresti essere coerente con le tue affermazioni” sono principi *interni*

³ John R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, Cambridge University Press 1969, tr.it. *Atti linguistici*, Boringhieri, Torino, 1976 ; *Expression and Meaning*, Cambridge, Cambridge University Press, 1979; *Intentionality*, Cambridge, Cambridge University Press, 1983, tr. it. *Della intenzionalità: un saggio di filosofia della mente*, Bompiani, Milano 1985; *The Construction of Social Reality*, New York, Basic Books, 1995, tr. it. *La costruzione della realtà sociale*, Comunità, Milano 1996.

alla nozione di asserzione. Non vi è alcun bisogno di un principio morale esterno per avere degli impegni rilevanti. L'impegno nei confronti della verità è insito nella struttura stessa dell'intenzionalità dell'asserzione.

II . Motivazione e direzione di adattamento

Fino ad ora ho presentato una spiegazione ridotta all'osso di come qualcuno possa creare degli impegni ed essere da questi motivato. In questo paragrafo voglio aggiungere alcuni ulteriori dettagli alla spiegazione. Francamente, la spiegazione fino a questo punto non mi sembra molto controversa, e nemmeno esaltante. Tuttavia devo dire che si incontra una fortissima resistenza. Perché? Gran parte della resistenza deriva dalla nostra tradizione filosofica secondo la quale una spiegazione del genere sarebbe impossibile. Secondo questa tradizione vi è una forte distinzione tra fatto e valore, tra "è" e "dovrebbe". La tradizione ha prodotto un numero infinito di libri concernenti il ruolo dei valori all'interno del mondo dei fatti e riguardanti le origini della normatività in tale mondo. La medesima tradizione manifesta un'ossessione dannosa nei confronti di qualcosa chiamato "etica" e "moralità", e gli autori sono di rado realmente interessati alle ragioni per l'azione, e sono troppo smaniosi di andare ad occuparsi del loro argomento preferito di etica. Considerano i fatti come non problematici e i valori come richiedenti una spiegazione. Ma se pensate alla questione dal punto di vista di quegli animali biologici sudati quali noi siamo, la normatività è pressoché dappertutto. Il mondo, infatti, consiste di fatti che sono ampiamente indipendenti da noi, ma se cominciamo a rappresentare quei fatti, con entrambe le direzioni di adattamento, avremo subito delle norme e queste norme sono vincolanti per gli agenti. L'intenzionalità ha una struttura normativa. Se un animale ha una credenza, la credenza è soggetta alle norme di verità, razionalità e coerenza. Se un animale ha delle intenzioni queste intenzioni possono avere successo o fallire. Se un animale ha delle percezioni, anch'esse possono avere successo o fallire nel fornire delle informazioni precise circa il mondo. E l'animale non può rimanere indifferente nei confronti della verità, del successo e della precisione, perché gli stati intenzionali in questione sono proprio gli stati di quegli animali. Se voi avete una credenza, io posso restare indifferente nei confronti della verità o della falsità della vostra credenza, ma se io ho una credenza, io non posso restare indifferente nello stesso modo,

perché è la mia credenza e la richiesta normativa di verità è insita nella credenza stessa. Dal punto di vista dell'animale, non c'è via d'uscita dalla normatività. La semplice rappresentazione di un *è* dà all'animale un *dovrebbe*.

Ciò che vi è di speciale negli animali umani non è la normatività, ma piuttosto l'abilità umana di creare, attraverso l'uso del linguaggio, un insieme *pubblico* di impegni. Gli essere umani fanno normalmente questo compiendo atti linguistici pubblici in cui il parlante impone intenzionalmente delle condizioni di soddisfazione su altre condizioni di soddisfazione. Questi atti linguistici sono resi possibili dall'esistenza di strutture istituzionali che il parlante usa per compiere degli atti linguistici dotati di significato e per comunicarli ad altri parlanti/ascoltatori. Usando questo apparato il parlante può assumersi degli impegni quando impone delle condizioni di soddisfazione su altre condizioni di soddisfazione. Di fatto non c'è modo di evitare di assumersi degli impegni. L'atto linguistico dell'asserzione è un impegno nei confronti della verità, l'atto linguistico della promessa è un impegno nei confronti di un'azione futura. Entrambi sorgono dal fatto che il parlante impone delle condizioni di soddisfazione su altre condizioni di soddisfazione. Gli atti linguistici impegnano il parlante nei confronti del secondo insieme di condizioni di soddisfazione. Nel caso di un'asserzione egli è impegnato nei confronti della verità dell'asserzione, nel caso della promessa è impegnato a compiere l'atto che ha promesso di compiere.

Una volta che la motivazione è creata, il suo riconoscimento fornisce una ragione per l'azione interna. È importante chiarire questo punto. L'accettazione di un qualsiasi motivatore esterno, anche se folle, può fornire all'agente una ragione per l'azione interna. Se io, in modo irrazionale, mi convinco che c'è una tigre nascosta sotto la mia scrivania, allora vuol dire che ho accettato l'esistenza di un pericolo e, di conseguenza, ho una ragione per agire, per quanto irrazionale possa essere la mia ragione. Il punto essenziale riguardante le ragioni per l'azione indipendenti dal desiderio è comunque che la loro accettazione è richiesta razionalmente come una questione di razionalità riconoscitiva, una volta che l'agente abbia creato intenzionalmente e liberamente la ragione in questione.

Consideriamo il caso che ho discusso in precedenza, dove affermo che sta piovendo. Ogni volta che faccio un'asserzione ho una ragione per parlare in maniera veritiera. Perché? Semplicemente perché un'asserzione è un impegno nei confronti della verità della proposizione espressa. Non c'è alcun divario tra l'esprimere un'asserzione e impegnare se

stessi nei confronti della sua verità. Ossia non vi sono due caratteristiche indipendenti dell'atto linguistico, la prima la costruzione di un enunciato e la seconda l'impegnare me stesso nei confronti della sua verità; c'è solo l'espressione dell'asserzione, che è *eo ipso* un impegno nei confronti della verità. Supponiamo che voi mi chiediate: "com'è il tempo fuori?" E io dica: "sta piovendo". *Per questo motivo* impegno me stesso nei confronti della verità della proposizione secondo la quale sta piovendo. Il mio impegno nei confronti della verità risulta ancora più evidente nei casi in cui mento. Se infatti non credo che stia piovendo, ma mento e dico "sta piovendo", il mio enunciato mi risulta intelligibile come una bugia proprio perché io comprendo che l'enunciato mi impegna nei confronti della verità di una proposizione che io non credo sia vera. E la bugia può avere successo come bugia proprio perché voi ritenete che io stia facendo un'asserzione e perciò impegnandomi nei confronti della verità della proposizione espressa. Una analoga considerazione può essere fatta per gli errori. Supponiamo che io non stia mentendo ma sia davvero in errore. Io ho detto in maniera sincera che stava piovendo, tuttavia non sta piovendo. In un caso del genere c'è ancora qualcosa di sbagliato per quanto riguarda il mio atto linguistico, ovvero esso è falso. Ma perché è sbagliato? Dopo tutto per ogni proposizione vera ce n'è una di falsa. Esso è sbagliato perché lo scopo dell'asserzione è di essere vera, e questo fallisce, perché essa è falsa. Quando faccio un'asserzione io mi impegno nei confronti della sua verità, e in questo caso il mio errore fa fallire il mio impegno.

Non c'è modo per il modello classico di spiegare questi semplici fatti. Il modello classico è spinto a sostenere che ci sono due fenomeni separati: l'espressione di un'asserzione e poi, esterno a questa, il principio per cui uno dovrebbe cercare di dire la verità. Che ragione ho per cercare di dire la verità quando esprimo un'asserzione o enunciato? Il teorico classico è spinto a sostenere che *io non ho nessuna ragione per il semplice fatto di esprimere un'asserzione*. La sola ragione che potrei avere sarebbe quella di avvertire che se mentissi potrebbero esserci delle cattive conseguenze, oppure l'adozione di un principio morale, logicamente indipendente dall'asserzione, in base al quale la falsità è sbagliata, o l'avvertire un'inclinazione a dire la verità, o l'avere qualche altra ragione esterna per esprimere quell'asserzione o enunciato. Secondo il modello classico tutte queste ragioni sono indipendenti dalla natura dell'esprimere enunciati in quanto tale. Io sto sostenendo, al contrario, che non c'è modo di spiegare che cos'è un enunciato se non spiegando che l'impegno nei confronti della verità è *interno* all'espressione dell'enunciato.

Ma perché l'impegno nei confronti della verità è interno all'espressione dell'enunciato? Perché non potremmo avere un diverso tipo di istituzione dell'esprimere enunciati, in cui noi esprimiamo enunciati ma non siamo impegnati nei confronti della loro verità? Qual è il grande problema dell'impegno? Ebbene, in un certo senso si possono compiere atti linguistici senza i loro normali impegni. Questo è ciò che accade nelle opere di narrativa. Nelle opere di narrativa nessuno ritiene l'autore responsabile della verità degli enunciati che egli ha prodotto nel testo. Comprendiamo questi casi come dei casi derivati da, e legati a, forme più fondamentali, in cui gli impegni riguardano le condizioni di verità degli enunciati effettivi. Per ripetere la domanda, perché allora? E la risposta è una conseguenza della natura del significato stesso. La ragione per cui sono impegnato nei confronti della verità dell'enunciato secondo il quale sta piovendo, quando dico che sta piovendo, è che, nell'esprimere l'enunciato che sta piovendo, io ho imposto intenzionalmente certe condizioni di soddisfazione a quell'enunciato. Assumendo che io non mi stia semplicemente esercitando sulla pronuncia, o facendo delle prove per una rappresentazione teatrale, o recitando una poesia, quando dico seriamente che sta piovendo sono impegnato nei confronti della verità della proposizione, poiché mi sono intenzionalmente impegnato nei confronti della verità di quell'enunciato quando ho imposto intenzionalmente le condizioni di soddisfazione che sta piovendo sulle condizioni di soddisfazione della mia intenzione-in-azione, quella intenzione-in-azione dovrebbe produrre i suoni "sta piovendo". E, lo ripeto, ciò che mi consente di fare questo in maniera pubblicamente accessibile è che io partecipo all'istituzione umana del linguaggio e degli atti linguistici.

Vorrei ora applicare alcune di queste considerazioni alla ragione pratica intesa in maniera più tradizionale. Molto spesso, nel caso della ragione pratica, si crea una ragione ora per compiere un atto nel futuro. Io credo che il solo modo per comprendere in che modo un'azione razionale volontaria possa creare delle ragioni per le azioni future è osservare la questione molto da vicino. Consideriamo dunque quei tipi di situazioni che accadono nella vita di tutti i giorni. Supponiamo che io vada in un bar e che ordini una birra. Supponiamo che io beva la birra e che arrivi il momento di pagare. Ora, la domanda è: dando per scontato il semplice fatto che io pensavo che il mio comportamento mi obbligasse a pagare la birra, devo avere anche una ragione indipendente da questo fatto, come il desiderio di pagare la birra, o qualche altro elemento appropriato del mio insieme motivazionale, per disporre di una ragione per pagare la birra? Cioè, per sapere se ho una ragione per pagare la birra, devo

prima analizzare il mio insieme motivazionale per vedere se contiene un qualche desiderio di pagare la birra, o vedere se dispongo di un qualche principio generale che riguarda il pagare la birra che ho bevuto? Mi sembra che la risposta sia no. In un caso del genere, ordinando la birra e bevendola una volta che mi è stata portata, io ho già creato intenzionalmente un impegno o un obbligo a pagarla e tali impegni e obblighi sono delle ragioni.

È assurdo che il modello classico non sia in grado di spiegare un caso così ovvio. Così come nel caso del dire la verità, il difensore del modello classico è spinto a dire che io ho una ragione per pagare la birra solo se posso individuare un desiderio rilevante nel mio “insieme motivazionale”. In opposizione a questo, intendo sostenere che in questa situazione ho creato una ragione per me stesso per pagare la birra, semplicemente ordinandola e bevendola.

Quali sono, esattamente, le caratteristiche formali della situazione che mi ha reso capace di creare una simile ragione? Quali sono, esattamente, le condizioni di verità dell'enunciato: l'agente A ha una ragione indipendente dal desiderio per compiere l'azione X nel futuro? Quale fatto che lo riguarda fa sì che egli abbia una tale ragione? Ebbene, un fatto che sarebbe sufficiente è il seguente: l'agente A ha *creato* un ragione indipendente dal desiderio per se stesso per compiere l'atto X nel futuro. Così ora la nostra domanda si riduce a: in che modo qualcuno crea una tale ragione? Ho già fornito una risposta a questa domanda come domanda logica circa le condizioni di soddisfazione, ma consideriamola ora “fenomenologicamente”. Che tipo di situazione appariva all'agente A quando ha ordinato la birra? Ebbene, se io sono l'agente, il modo in cui la situazione mi appare è questo: in questo momento sto compiendo l'atto che fa sì che il barista mi porti una birra sapendo che io ho l'obbligo di pagarla se lui me la porta. Ma se questa è l'intenzione, allora, in questa azione effettiva, se il barista porta la birra, io ho reso possibile il fatto di avere un obbligo, e perciò una ragione, che sarà per me una ragione per agire nel futuro, e la ragione che ho creato sarà indipendente dai miei altri desideri futuri. In questo caso, una condizione sufficiente perché un atto crei una ragione per me è il fatto stesso che io voglio che esso crei una ragione per me.

Il meccanismo formale con il quale ho creato l'obbligo è esattamente parallelo al meccanismo formale con cui ho creato un impegno nel caso dell'espressione di un enunciato. In questo caso, tuttavia, ho imposto delle condizioni di soddisfazione al mio

enunciato, il quale aveva una direzione di adattamento verso l'alto. Io mi sono assunto l'obbligo di fare qualcosa. Questo è difficile da vedere, perché non l'ho fatto esplicitamente nell'enunciato. Ho semplicemente detto "portami una birra", e questo enunciato ha le condizioni di soddisfazione con direzione di adattamento verso l'alto, ovvero che l'ascoltatore dovrebbe portarmi una birra. Ma la comprensione totale della situazione, che avremo occasione di esplorare in dettaglio quando considereremo la promessa, è che io ho anche imposto delle condizioni di soddisfazione su me stesso, sul mio comportamento futuro. E le ho imposte sotto forma di un obbligo condizionale. Gli obblighi hanno una direzione di adattamento verso l'alto, o mondo-a-obbligo. L'obbligo è soddisfatto o adempiuto soltanto se il mondo cambia, solitamente nella forma del comportamento della persona che ha quell'obbligo, per adattarsi al contenuto dell'obbligo. Gli obblighi sono perciò una specie di motivatori esterni. Normalmente la loro esistenza è epistemicamente oggettiva, sebbene, dal momento che sono sempre creati da esseri umani ed esistono soltanto in relazione alle attitudini degli esseri umani, essi siano ontologicamente soggettivi. E, come abbiamo avuto occasione di osservare più volte, la soggettività ontologica non implica la soggettività epistemica. Può essere un semplice dato di fatto che io mi impegni in un obbligo, sebbene la creazione e l'esistenza dell'obbligo siano relativi all'osservatore.

La presupposizione della libertà dell'agente è cruciale per il caso che ho descritto. Dal punto di vista della prima persona, assumendo liberamente di creare di una ragione per me stesso, ho già manifestato un desiderio che quella sia una ragione per me. Ho già legato la mia volontà nel futuro attraverso il libero esercizio della mia volontà nel presente. Alla fine tutte queste domande devono avere delle risposte banali. Perché essa è una ragione? Perché l'ho creata come ragione. Perché è una ragione *per me*? Perché l'ho liberamente creata come ragione per me.

Nella discussione sul divario esposta nei capitoli 1 e 3, abbiamo osservato che tutte le ragioni effettive sono create dall'agente. Ma la peculiarità della creazione di ragioni indipendenti dal desiderio per le azioni future è che io, in questo momento, tramite l'esercizio di una ragione effettiva, ho creato una ragione potenzialmente effettiva per agire nel futuro. La tradizione filosofica ha il problema esattamente opposto. Il problema non è "come potrebbero esserci per me delle ragioni indipendenti dal desiderio?", ma piuttosto "come potrebbe qualsiasi cosa essere una ragione di qualsiasi tipo per me che non l'ho creata come una ragione per me, incluse le ragioni indipendenti dal desiderio?". Nel

compiere un'azione volontaria c'è un divario tra le cause e l'effettivo compimento dell'azione, e questo divario si presenta semplicemente quando compio l'azione; in questo caso l'esecuzione dell'azione è essa stessa la creazione di una ragione per un'azione conseguente.

Per quanto riguarda la motivazione, nei casi che ho descritto *la ragione può essere il fondamento del desiderio e non viceversa*. In italiano corrente la corretta descrizione di questo caso è “voglio pagare per questa cosa perché ho l'obbligo di farlo”. E la connessione tra ragione, razionalità e desiderio è la seguente: il riconoscimento di qualcosa come un obbligo vincolante è già il riconoscimento di qualcosa la cui ontologia è quella di un motivatore esterno, ossia di un'entità che ha una direzione di adattamento verso l'alto. Riconoscere la validità di una tale entità è già riconoscere una ragione per agire. E il riconoscimento di qualcosa in quanto ragione per agire è già il riconoscimento di quella cosa in quanto ragione per desiderare di compiere l'azione.

III . La soluzione di Kant al problema della motivazione

Immanuel Kant, nella *Fondazione della metafisica dei costumi*⁴, affrontava un problema che è formalmente simile a quello che sto discutendo. Il mio problema è: come possono le ragioni indipendenti dal desiderio motivare realmente le azioni, se ogni azione è l'espressione del desiderio di compiere quella particolare azione? Kant enuncia il suo problema nella forma seguente: “come può la ragione pura essere pratica?” E lo spiega sostenendo che esso esprime la domanda del perché possiamo provare *interesse* nei confronti dell'imperativo categorico. Un interesse è ciò in virtù di cui una ragione diventa pratica, cioè, diventa una causa che determina la volontà di agire. Mi sembra che la risposta di Kant a questa domanda sia inadeguata. Così egli afferma: "Per volere ciò che soltanto la ragione stabilisce come dovere per un essere ragionevole affetto sensibilmente, è certamente richiesta una capacità della ragione di *ispirare un sentimento di piacere* o di soddisfazione per il dovere compiuto, quindi una causalità che determini la sensibilità in conformità ai suoi principi" (p. 104). Secondo la prospettiva di Kant, la ragione pura deve dunque causare

⁴ Immanuel Kant, *Grundlegung der Metaphysik der Sitten*, 1785, tr. it. *Fondazione della metafisica dei costumi*, Laterza, Roma-Bari 1993, p. 104.

un sentimento di piacere, ed è solo a causa di quel sentimento di piacere che noi siamo realmente capaci di agire in accordo con i precetti della ragione pura. Kant ammette che ci risulta totalmente inintelligibile il modo in cui la ragione pura possa anche solo causare questo sentimento di piacere, perché noi possiamo solo scoprire le relazioni di causa ed effetto tra gli oggetti di esperienza, e la ragione pura non è un oggetto di esperienza.

Penso che questo sia un cattivo argomento. La tesi di Kant è che noi non potremmo agire sulla base di ragioni per l'azione indipendenti dal desiderio a meno che, in un modo o nell'altro, non provassimo un "sentimento di piacere" nel fare questo. Penso che Kant non comprendesse la nozione di direzione di adattamento. Ovvero, io penso che noi possiamo compiere molte azioni nelle quali non c'è alcun "sentimento di piacere", ma c'è soltanto il riconoscimento che noi abbiamo una ragione valida per compierle. Non devo provare un "sentimento di piacere" quando mi faccio trapanare un dente più di quanto non devo provare un sentimento di piacere quando mantengo le mie promesse. Potrei ricavare soddisfazione dal farmi trapanare il dente e dal mantenimento delle mie promesse, ma non è logicamente necessario che io abbia un sentimento del genere perché io mi faccia trapanare un dente o mantenga le promesse. *Secondo la prospettiva che sto presentando, il riconoscimento della validità della ragione è sufficiente per motivare l'azione.* Non c'è bisogno di alcun ulteriore piacere, desiderio o soddisfazione. La motivazione per compiere l'azione corrisponde esattamente alla motivazione per voler compiere l'azione.

Questo è un punto assolutamente cruciale, sia per l'argomento di Kant che per quello di questo libro e, per la verità, per il dibattito sul modello classico in generale. Kant, sebbene attacchi il modello classico in vari modi, accetta una delle sue peggiori caratteristiche. Kant assume che io non posso compiere un'azione intenzionalmente e volontariamente qui ed ora, a meno che io non abbia un "sentimento di piacere", qui ed ora, nel compiere quell'azione. Se qualsiasi azione è davvero compiuta per soddisfare un desiderio e se qualsiasi azione è essa stessa l'espressione del desiderio di compiere quell'azione, allora ci deve essere una qualche soddisfazione del desiderio nel compiere qualsiasi azione. Ma questo è un ginepraio di confusioni e cercherò ora di districarlo. Per prima cosa consideriamo i casi in cui l'azione è compiuta per soddisfare un desiderio. Io mi faccio trapanare il dente per far sì che esso ritorni sano. E me lo faccio trapanare perché voglio, in quel momento, farmelo trapanare. Ma da questo non segue in nessun senso la necessità di un "sentimento di piacere" nella mia azione intenzionale. Il mio desiderio

primario di far sì che mio dente ritorni sano può motivare il desiderio secondario di farmelo trapanare, che a sua volta motiva l'azione. Ma il piacere o la soddisfazione che deriva dal farmi riparare il dente non comporta necessariamente l'attività di farmelo trapanare, né il bisogno che questo accada. Questo è un caso in cui ho una ragione dipendente dal desiderio per desiderare qualcosa, ma il modo in cui la ragione dipendente dal desiderio fonda il desiderio secondario è esattamente lo stesso modo in cui una ragione indipendente dal desiderio fonda un desiderio secondario. Il mio desiderio di mantenere una promessa deriva dal fatto indipendente dal desiderio che io ho fatto una promessa, e per questo ho un obbligo. Ma il mio derivare un sentimento di piacere dal mantenere una promessa, al fine di compiere intenzionalmente l'azione di mantenere la mia promessa, non è più necessario di quanto lo sia il fatto che io provi un sentimento di piacere dal farmi trapanare un dente per poter soddisfare il mio desiderio primario di farmi riparare il dente. L'errore di Kant rende esplicito un errore che, negli autori appartenenti alla tradizione classica, è soltanto implicito. Se ogni azione è l'espressione di un desiderio di compiere quell'azione, ed ogni azione che ha successo ha come risultato la soddisfazione del desiderio, allora sembra che la sola cosa che possa motivare l'azione sia la soddisfazione del desiderio, cioè, un sentimento di piacere. Ma questa è una fallacia. Dal fatto che ogni azione è effettivamente l'espressione di un desiderio di compiere quell'azione, non segue che ogni azione sia fatta con il *proposito* di soddisfare un desiderio, né segue che le azioni possano essere motivate solamente dalla soddisfazione del desiderio, nel senso di un sentimento di piacere.

IV. La promessa come caso speciale

Le discussioni di questi problemi solitamente dedicano molto spazio al caso della promessa, ma io sto cercando di evidenziare che il fenomeno delle ragioni indipendenti dal desiderio create dall'agente è pervasivo. Non si potrebbe comprendere la vita sociale senza di esso, e il promettere è solo un caso speciale e un tipo di caso particolarmente chiaro. La storia dei dibattiti sulla promessa è tuttavia rivelatrice e sarò in grado di spiegare meglio quanto sto sostenendo se spiegherò l'obbligo di mantenere una promessa e se esporrò alcuni degli errori tipici. La domanda è: che ragione abbiamo per mantenere una promessa? E la risposta ovvia a questo è: le promesse sono, per definizione, delle creazioni di obblighi; e gli obblighi sono, per definizione, delle ragioni per l'azione. E una domanda segue immediatamente: qual è l'*origine* dell'obbligo di mantenere una promessa?

Non c'è alcun modo in cui il modello classico possa spiegare il fatto che l'obbligo di mantenere una promessa è interno all'atto di promettere, proprio come l'impegno nei confronti del dire la verità è interno all'atto del fare asserzioni. Cioè, promettere è, per definizione, assumere un obbligo a fare qualcosa. La tradizione è spinta a negare questo fatto, ma per negarlo i difensori del modello classico, normalmente, sono spinti a dire cose strane e, io credo, sbagliate. In questo paragrafo offrirò un breve elenco degli errori più comuni che ho incontrato.

Ci sono tre posizioni comuni, a mio avviso errate, delle quali ci si può sbarazzare velocemente. La prima è supporre che vi sia uno speciale obbligo *morale* per mantenere una promessa. Al contrario, se ci pensate vedrete che non c'è alcuna connessione speciale tra il promettere e la moralità, strettamente intesa. Se io prometto di venire al vostro ricevimento, per esempio, questo costituisce un obbligo sociale. Che questo sia anche un obbligo morale dipende dalla natura del caso, ma per la gran parte dei ricevimenti a cui vado non sarebbe un obbligo morale. Spesso facciamo delle promesse riguardanti alcune questioni morali serie, ma non c'è niente riguardo al promettere in quanto tale che implichi che ogni promessa coinvolga un'istanza morale. Non c'è nulla nel promettere in quanto tale che garantisca che ogni obbligo di mantenere una promessa sia abbastanza serio da essere considerato un obbligo morale. Si possono fare delle promesse riguardo a questioni che sono banali, da un punto di vista morale.

Un secondo errore è supporre che se si promette di fare qualcosa di malvagio, non vi sia alcun obbligo a mantenere la promessa. Ma questo è ovviamente sbagliato. Il modo corretto per descrivere tali casi è quello di dire che si ha invece un obbligo a mantenere la promessa ma che questo è sovrastato dalla natura maligna dell'atto promesso. Questo punto può essere dimostrato con il metodo dell'accordo e della differenza: c'è una differenza tra la persona che ha promesso di compiere l'atto e la persona che non lo ha promesso. La persona che ha fatto la promessa ha una ragione che la persona che non ha fatto la promessa non ha.¹

Il terzo, e credo il peggiore, degli errori è supporre che un obbligo a mantenere una promessa sia solo un obbligo *prima facie*, rispetto ad un obbligo esplicito. Questa tesi fu formulata (da Sir David Ross)² per cercare di venire a capo del fatto che gli obblighi, comunemente, sono in conflitto tra loro e spesso non possono essere soddisfatti tutti. Quando un obbligo A sovrasta l'obbligo B, dice Ross, B è solo un obbligo *prima facie*, non un obbligo esplicito. Ho sostenuto in dettaglio altrove³ che questa teoria è confusa e non ripeterò gli argomenti in questa sede se non per dire che quando B è superato da qualche obbligo più importante, questo non dimostra che B non fosse un obbligo a tutti gli effetti, incondizionato, ecc. Non lo si può superare, in primo luogo, se non c'è, in realtà, nulla con cui superarlo. "Prima facie" è un enunciato modificatore di tipo epistemico, non un predicato di tipi di obbligazioni, e non potrebbe essere in alcun modo un termine appropriato per descrivere il fenomeno degli obblighi in conflitto tra loro, in cui uno è superato o sovrastato dall'altro. La teoria delle "obbligazioni *prima facie*" è una cosa peggiore di una cattiva teoria filosofica: è cattiva grammatica.

Io credo che i seguenti siano i più seri errori comuni circa l'obbligo di mantenere una promessa, e derivano tutti, se pure in modi differenti, dall'accettare il modello classico:

Errore numero 1: l'obbligo di mantenere una promessa è prudenziale. La ragione per mantenere una promessa è che se non la manterrò, quando farò delle promesse in futuro non mi si crederà più.

¹ Nella legge, un contratto per fare qualcosa di illegale è considerato nullo e vuoto e non può essere fatto valere in tribunale. Questo non perché non vi fosse un contratto, ma perché la legge lo rende nullo.

² W.D.Ross, *The Right and the Good*, Oxford, Oxford University Press, 1930, p.28.

³ John R. Searle, "Prima Facie Obligations", in Zac van Straaten (ed.), *Philosophical Subjects: Essays Presented to P.F.Strawson*, Oxford, Oxford University Press, 1980, pp. 238-260.

Come è noto, Hume accolse questa tesi. Ma essa è soggetta a un'obiezione decisiva e altrettanto nota: secondo questa spiegazione, nei casi in cui nessun essere vivente conoscesse la mia promessa, io non avrei nessun obbligo a mantenerla. Secondo tale prospettiva, la promessa fatta in letto di morte, fatta dal figlio in privato al padre morente, non comporterebbe nessun obbligo perché il figlio non ha bisogno di dire a nessuno la promessa.

Inoltre, perché non mi si dovrebbe credere in futuro? Solo perché avevo assunto un obbligo e non sono riuscito a mantenerlo. Il fallimento nel mantenere gli obblighi come fondamento per la sfiducia è piuttosto diverso dal semplice fatto di disattendere le aspettative. Per esempio, Kant, come è noto, andava a passeggiare in modo talmente regolarmente che i suoi vicini potevano regolare i loro orologi in base alle sue passeggiate. Se avesse un giorno fallito nel fare la passeggiata all'ora usuale, egli avrebbe disatteso delle aspettative, ma non avrebbe ispirato sfiducia allo stesso modo della persona che rinnega i suoi obblighi. Nel caso della promessa la sfiducia sorge non semplicemente per il fallimento di un'aspettativa, bensì dal fatto che colui che ha fatto la promessa aveva dato la sua parola.

Errore numero 2: l'obbligo a mantenere una promessa deriva dall'accettazione di un principio morale il cui contenuto è che una persona dovrebbe mantenere le sue promesse. Senza tale accettazione l'agente non ha ragioni, eccetto forse quelle prudenziali, per mantenere una promessa.

L'errore qui è uguale a quello che abbiamo trovato nel caso dell'impegno nei confronti della verità quando si fa un'asserzione. Il modello classico cerca di rendere l'obbligo della promessa esterno all'atto del promettere, ma in questo modo diventa impossibile spiegare che cos'è una promessa, così come diventerebbe impossibile spiegare che cos'è un'asserzione se si prova a considerare la relazione tra l'asserire e l'impegnare se stessi alla verità dell'asserzione come una relazione puramente esterna. Cioè, la risposta decisiva a questa obiezione è far vedere che le relazioni tra il promettere e gli obblighi sono interne. Per definizione, una promessa è un atto che riguarda l'assunzione di un obbligo. È impossibile spiegare che cos'è una promessa se non nei termini dell'assunzione di un obbligo.

Proprio come abbiamo osservato nel caso dell'espressione di enunciati o asserzioni, dove l'impegno nei confronti della verità si rivelava in modo più evidente quando la persona mentiva deliberatamente, così nel caso del promettere possiamo mostrare che l'obbligo è interno all'atto del promettere in modo più evidente quando una persona fa una promessa non sincera. Supponiamo che io faccia una promessa non sincera, una promessa che non ho alcuna intenzione di mantenere. In questo caso il mio atto di inganno è per me pienamente intelligibile, e in seguito può essere considerato da colui al quale ho fatto la promessa come un atto disonesto, esattamente perché si capisce che quando ho fatto la promessa io stavo impegnando me stesso, assumendo un obbligo, a fare la cosa che avevo promesso di fare. Quando faccio una promessa io non azzardo un'ipotesi, né faccio una predizione su che cosa accadrà nel futuro; piuttosto io vincolo la mia volontà a ciò che farò in futuro. *La mia promessa disonesta risulta intelligibile a me stesso come una promessa in cui mi assumo un obbligo senza alcuna intenzione di mantenere l'obbligo che ho sottoscritto.*

Errore numero 3 (questa è una variante più sofisticata del numero 2): se gli obblighi fossero davvero interni alla promessa allora l'obbligo di mantenere una promessa dovrebbe derivare dall'istituzione del promettere. Il fatto che qualcuno ha fatto una promessa è un fatto istituzionale, e qualsiasi obbligo dovrebbe derivare dall'istituzione. Ma, allora, che cosa impedisce che ogni istituzione abbia lo stesso *status*? La schiavitù è un'istituzione come il promettere. Così se fosse vera la prospettiva secondo la quale le promesse creano delle ragioni indipendenti dal desiderio, allora lo schiavo avrebbe un obbligo così come colui che fa una promessa, il che è assurdo. Quindi, poiché la teoria della promessa indipendente dal desiderio, porta a dei risultati assurdi essa deve essere falsa. Il modo corretto di vedere le cose è osservare che l'istituzione è in verità il fondamento dell'obbligo ma solamente perché, indipendentemente dall'istituzione, noi accettiamo il principio che una persona dovrebbe mantenere le proprie promesse. Senza l'approvazione di un'istituzione o di qualcosa che la supporti o in qualche modo la valuti favorevolmente, non vi sarebbe nessun obbligo della promessa. Noi siamo abitualmente portati a mantenere le nostre promesse e ad adottare così un'atteggiamento favorevole verso l'istituzione; per questo non riusciamo a notare che la nostra approvazione dell'istituzione è la fonte essenziale dell'obbligo. In quanto istituzioni, la promessa e la schiavitù sono coerenti; la sola differenza, per quanto riguarda la nostra discussione, è che ci capita di pensare che uno è buono mentre l'altra è

cattiva. Ma l'obbligo non è interno all'atto di promettere, deriva esternamente dall'attitudine che abbiamo verso l'atto del promettere. Il solo modo in cui potrebbe essere creato l'obbligo del promettere è che noi accettassimo il principio “non devi infrangere le tue promesse.

Questa obiezione incapsula la teoria del modello classico in questo problema. La risposta più semplice è la seguente: *l'obbligo di mantenere una promessa non deriva dall'istituzione del promettere*. Quando faccio una promessa l'istituzione del promettere è semplicemente il *veicolo*, lo strumento che uso per creare una ragione. L'obbligo di mantenere una promessa deriva dal fatto che nel promettere io creo liberamente e volontariamente una ragione per me stesso. Il libero esercizio della volontà può vincolare la volontà, e questo è un passaggio logico che non ha nulla a che vedere con le “istituzioni” o le “attitudini morali” o gli “enunciati di valore”. Questo è il motivo per cui lo schiavo non ha alcuna ragione per obbedire al padrone, ad eccezione delle ragioni prudenziali. Egli non ha vincolato la sua volontà con un esercizio della sua libertà. Visto dall'esterno, lo schiavo potrebbe apparire esattamente come un lavoratore a contratto. Potrebbero persino avere entrambi lo stesso salario. Ma dall'interno la questione è piuttosto differente. Il lavoratore a contratto ha creato una ragione per se stesso che lo schiavo non ha creato. Pensare che l'obbligo del promettere derivi dall'istituzione del promettere è un errore come il pensare che gli obblighi che assumo quando parlo in inglese debbano derivare dall'istituzione dell'inglese: a meno che non pensi che l'inglese sia in qualche modo una cosa buona, io non ho nessun obbligo quando lo parlo. Secondo il modello classico, l'obbligo di mantenere una promessa è sempre qualcosa di esterno alla promessa stessa. Se io ho un obbligo di mantenere una promessa, ciò può essere solo perché io penso (a) che l'istituzione del promettere sia una cosa buona o (b) adottato un principio morale il cui contenuto è che una persona dovrebbe mantenere le proprie promesse. C'è una confutazione piuttosto semplice di entrambe queste prospettive: esse hanno come conseguenza che in assenza di queste condizioni non ci sarebbe nessun obbligo di mantenere una promessa. Così, per chi non pensi che l'istituzione del promettere sia una buona cosa, o per chi non adotti un principio morale secondo il quale una persona dovrebbe mantenere le proprie promesse, non c'è nessuna ragione per mantenere una promessa. Credo che questo sia assurdo ed ho evidenziato tale assurdità in diversi punti di questo libro.

Errore numero 4: ci sono realmente due diversi sensi di tutte queste parole: “promessa”, “obbligo” ecc., un senso descrittivo e un senso valutativo. Secondo il senso descrittivo, quando usiamo queste parole, noi riportiamo semplicemente dei fatti, non adottiamo in effetti nessuna ragione per l’azione. Quando le usiamo in senso valutativo, vi è molto di più del semplice enunciare dei fatti, poiché in questi casi noi dobbiamo esprimere dei giudizi morali, e questi giudizi morali non possono mai derivare dai fatti in sé. Così, in realtà, c’è un’ambiguità sistematica nell’intera discussione. L’ambiguità è tra i significati descrittivi e valutativi delle parole.

Sarò breve nel rispondere all’errore numero 4. Non vi sono due diversi sensi per queste parole più di quanto non vi siano due sensi per “cane”, “gatto”, “casa” o “albero”. Naturalmente si possono sempre usare le parole in un modo che non implica i normali impegni. Invece di dire “quella è una casa”, posso dire “quella è ciò che essi chiamano ‘una casa’”, e in questo caso io non impegno me stesso in un modo o nell’altro su che cosa sia effettivamente una casa (sebbene io impegni me stesso nei confronti delle persone che la chiamano in quel modo). Ora, allo stesso modo, se dico “egli ha fatto una promessa” o “egli ha assunto un obbligo” io posso mettere le virgolette alle parole “promessa” e “obbligo” e rimuovere così l’impegno sostenuto dal significato letterale delle parole. Ma questa possibilità non mostra che ci sono due sensi per ciascuna di queste parole o che ci sia una qualche ambiguità nel loro uso letterale. Il significato letterale di “promessa” è tale che qualcuno che abbia fatto una promessa ha in questo modo assunto un obbligo a fare qualcosa. Cercare di postulare dei sensi ulteriori per queste parole significa evadere da questi argomenti.

V. Generalizzare la spiegazione: il ruolo sociale delle ragioni indipendenti dal desiderio

Finora in questo capitolo ho cercato di descrivere quella che chiamerei la struttura atomica della creazione di ragioni per l’azione indipendenti dal desiderio e ho discusso alcune caratteristiche specifiche delle asserzioni e delle promesse, con una particolare enfasi nel criticare la tradizione filosofica e la sua discussione a riguardante l’istituzione del

promettere. Ho anche brevemente discusso il “livello fenomenologico” delle ragioni per l’azione indipendenti dal desiderio, in cui una persona agisce sulla base della comprensione che l’azione di una persona creerà una ragione per se stessa per fare qualcosa nel futuro. Ora voglio provare ad esporre una spiegazione più generale del ruolo delle ragioni indipendenti dal desiderio nella vita sociale in generale, a un livello più ampio di quello della struttura atomica. Vorrei, tra le altre cose, spiegare perché la creazione di ragioni indipendenti dal desiderio da parte di sé liberi e razionali, in possesso di un linguaggio e operanti all’interno di strutture istituzionali, sia così pervasiva. Questo è ciò che accade quando ci si sposa, si ordina una birra in un bar, si acquista una casa, ci si iscrive ad un corso universitario o si prende un appuntamento con il proprio dentista. In questi casi si invoca una struttura istituzionale in modo tale da creare un ragione per se stessi per fare qualcosa nel futuro indipendentemente dal fatto che nel futuro si abbia un desiderio di fare quella cosa. E in tali casi questa costituisce una ragione per voi perché voi l’avete volontariamente creata come una ragione per voi.

Una spiegazione generale del ruolo delle ragioni nella razionalità pratica implica perlomeno la comprensione delle cinque seguenti caratteristiche: (1) *libertà*; (2) *temporalità*, (3) *il sé e con esso il punto di vista della prima persona*; (4) *il linguaggio e altre strutture istituzionali*; (5) *la razionalità*. Consideriamo ognuna di queste in ordine.

Libertà

Ho già sostenuto che la razionalità e il presupposto della libertà sono coestensivi. Essi non sono la stessa cosa ma le azioni sono valutabili razionalmente se e solo se le azioni sono libere. La ragione di questa connessione è la seguente: *la razionalità deve essere capace di porre una differenza*. La razionalità è possibile solo dove c’è una scelta genuina tra i vari corsi di azione razionali e irrazionali. Se l’atto è completamente determinato allora la razionalità non può porre alcuna differenza. Non entra nemmeno in gioco. La persona il cui atto è interamente causato da credenze e da desideri nel modo teorizzato dal modello classico agisce in maniera coercitiva al di fuori dell’ambito della razionalità in generale. Ma la persona che liberamente *agisce sulla base* di quelle stesse credenze e desideri, colui che liberamente li trasforma in ragioni *effettive*, agisce all’interno del regno della razionalità. La libertà di azione, il divario e l’applicabilità della razionalità sono coestensivi.

Agendo liberamente, io posso, imponendo condizioni di soddisfazione su condizioni di soddisfazione, creare una ragione che sarà una ragione per me per fare qualcosa nel futuro, indipendentemente dal fatto che avrò voglia di farla quando arriverà quel momento. L'abilità di vincolare la volontà ora può creare una ragione per l'atto futuro solamente perché è una manifestazione della libertà.

Temporalità

Gli enunciati della ragione teorica sono atemporalmente nello stesso modo in cui gli enunciati della ragione pratica sono intrinsecamente temporali. L'enunciato "farò l'atto A perché voglio che si dia il caso che B" è essenzialmente riferito al futuro così come "l'ipotesi H è dimostrata dall'evidenza E" non è affatto essenzialmente temporalizzato. È senza tempo, sebbene, certamente, in ambiti particolari, possa fare riferimento a situazioni storiche particolari. Per gli animali non umani esistono effettivamente solo ragioni immediate, dal momento che senza il linguaggio non si può organizzare il tempo.

Il sé e il punto di vista della prima persona

Nei casi che considereremo, è essenziale notare che stiamo esaminando la struttura logica del comportamento di sé razionali impegnati nel creare ragioni per se stessi. Nessun punto di vista esterno o in terza persona può spiegare i processi attraverso i quali un agente libero può creare una ragione ora che lo vincolerà nel futuro, indipendentemente da come si potrà sentire nel futuro.

Il linguaggio e altre strutture istituzionali

Per creare delle ragioni indipendenti dal desiderio un agente deve possedere il linguaggio. Si possono immaginare degli esseri primitivi pre-linguistici che impongono condizioni di soddisfazione su condizioni di soddisfazione. Ma la creazione sistematica di tali ragioni, e la loro comunicazione ad altre persone, richiede la presenza di strumenti simbolici convenzionali con le caratteristiche tipiche del linguaggio umano. Le relazioni sociali richiedono Inoltre che noi siamo capaci di rappresentare le relazioni deontiche implicite nella creazione di ragioni per l'azione indipendenti dal desiderio e che abbiamo bisogno di un linguaggio per ordinare il tempo nel modo richiesto. Dobbiamo cioè disporre di modi di rappresentazione del fatto che l'azione presente crea una ragione per un'azione

futura, e dobbiamo disporre di modi linguistici di rappresentazione delle relazioni temporali e deontiche in questione.

In aggiunta al linguaggio strettamente inteso, cioè, in aggiunta ad atti linguistici come l'asserzione o il fare una promessa, ci sono anche strutture istituzionali extralinguistiche che funzionano nella creazione di ragioni indipendenti dal desiderio. Così, per esempio, solo se una società ha l'istituzione della proprietà ci possono essere ragioni indipendenti dal desiderio che implicano la proprietà, e solo se una società ha l'istituzione del matrimonio ci possono essere ragioni indipendenti dal desiderio che implicano l'istituzione del matrimonio. Il punto, comunque, che deve essere enfatizzato più volte, è che la ragione non deriva dall'istituzione, piuttosto l'istituzione fornisce l'intelaiatura, la struttura, all'interno della quale si crea la ragione. La ragione deriva dal fatto che l'agente vincola la sua volontà attraverso un atto libero e volontario.

Razionalità

Affinché la pratica del creare ragioni indipendenti dal desiderio possa essere socialmente effettiva, essa deve essere effettiva in virtù della razionalità degli agenti coinvolti. È solo perché sono un'agente razionale che posso riconoscere il fatto che il mio comportamento precedente ha creato delle ragioni per il mio comportamento presente.

La combinazione di questi cinque elementi

Proviamo a mettere insieme questi punti in una spiegazione generale. Tanto per iniziare, come possiamo organizzare il tempo? La risposta ovvia è che noi facciamo delle cose ora che faranno accadere delle cose nel futuro così che esse non sarebbero accadute se non avessimo agito in quel modo ora. Questo è il motivo per cui mettiamo la sveglia per le 6.00 di mattina. Noi sappiamo di avere una ragione per alzarci alle 6:00, ma sappiamo anche che alle 6:00 non saremo capaci di agire sulla base di questa ragione perché saremo ancora addormentati. Così, regolando la sveglia ora, creiamo la possibilità di agire sulla base di una ragione nel futuro. Ma supponiamo che io non abbia una sveglia e debba farmi svegliare da qualcun altro. Qual è la differenza tra mettere la sveglia alle 6:00 e chiedere a qualcuno di svegliarmi alle 6:00, per esempio? In entrambi i casi io faccio qualcosa ora per far sì che io mi svegli alle 6:00 domani. La differenza è che nel caso della sveglia vengono create solo delle cause, mentre nel secondo caso vengono create delle nuove ragioni per l'azione. In che

modo? Ebbene ci sono diversi tipi di casi. Se non ho fiducia nella persona in questione io potrei dire “se mi svegli alle 6:00 ti do cinque dollari”. In quel caso io faccio una promessa, una promessa condizionale di dare cinque dollari all’altra persona e, se egli accetta l’offerta, promette di svegliarmi sotto la condizione che io gli paghi cinque dollari. Questo è tipico dei contratti. Ogni parte fa una promessa condizionale, condizionale sul ricevere un beneficio dall’altra parte.

Nel caso più realistico io strappo a lui la promessa di svegliarmi. Io dico “per favore svegliami alle 6:00” e lui dice “va bene”. In quel contesto egli ha fatto una promessa incondizionata ed ha creato una ragione indipendente dal desiderio.

In un terzo tipo di caso, non c’è bisogno di fare nessuna promessa. Supponiamo che io non abbia nessuna fiducia in quella persona, ma so che egli fa colazione ogni giorno alle 6:00. Io semplicemente sistemo il cibo della sua colazione in modo tale che egli non possa prenderlo senza svegliarmi. Lo porto nella mia camera e chiudo la porta a chiave, per esempio. Per fare colazione egli deve bussare alla mia porta e svegliarmi. Ora anche questo terzo tipo di caso crea una ragione per svegliarmi, ma questa è una ragione dipendente dal desiderio o prudenziale. Egli deve fare questo ragionamento: “io voglio la colazione, non posso avere la colazione senza svegliarlo, quindi lo sveglio”. Tutti e tre questi metodi possono funzionare ugualmente bene, ma vorrei richiamare l’attenzione su quanto sia bizzarro il terzo caso. Se il solo modo in cui possiamo ottenere cooperazione da altre persone fosse quello di metterle nella condizione in cui esse, indipendentemente da noi, vogliono fare ciò che noi vogliamo che essi facciano, gran parte delle forme della vita sociale umana sarebbe impossibile. *Per poter organizzare il tempo su una base sociale è necessario che noi creiamo dei meccanismi per giustificare delle aspettative ragionevoli circa il comportamento futuro dei membri della comunità, inclusi noi stessi.* Se avessimo soltanto dei desideri, come le scimmie di Köhler, noi non saremmo mai capaci di organizzare il tempo così da consentirci di organizzare il nostro comportamento, e di coordinarci con gli altri sé. Per organizzare e coordinare il nostro comportamento, noi dobbiamo creare una classe di entità che abbia la stessa struttura logica dei desideri, ma che sia indipendente dal desiderio. In breve, dobbiamo creare una classe di motivatori esterni che fornisca una ragione per l’azione – cioè, un contenuto proposizionale con una direzione di adattamento verso l’alto, e l’agente come soggetto. Il solo modo in cui tali entità possono

essere vincolate a dei sé razionali è, esattamente, che i sé razionali le creino liberamente in quanto vincolanti loro stessi.

Volgiamoci ora al ruolo del linguaggio e ad altre strutture istituzionali. Ci sono molte caratteristiche dei fatti istituzionali che richiedono un'analisi; altrove ho tentato di fornire un'analisi di queste e non la ripeterò in questa sede.⁴ C'è tuttavia una caratteristica che risulta essenziale per la discussione presente. Nel caso dei fatti istituzionali, la normale relazione tra intenzionalità e ontologia è invertita. Nel caso normale, ciò che è risulta logicamente precedente a ciò che *sembra essere*. Così, noi comprendiamo che quell'oggetto sembra essere pesante perché comprendiamo che cosa significa per un oggetto essere pesante. Ma nel caso della realtà istituzionale, l'ontologia deriva dall'intenzionalità. Affinché un certo tipo di cosa sia denaro, è necessario che le persone pensino che esso sia denaro. Se un numero sufficiente di loro pensa che esso sia denaro, ha altri atteggiamenti appropriati e agisce in maniera appropriata, e se quel tipo di cosa soddisfa tutte le altre condizioni stabilite dai loro atteggiamenti, come il non essere contraffatto, allora quella cosa è denaro. Se tutti noi pensiamo che un certo tipo di cosa sia del denaro e concordiamo nell'usarla, nel considerarla, nel trattarla come denaro, allora esso è denaro. In questo caso "sembra" è precedente a "è". Non si può esagerare sull'importanza data a questo fenomeno. I suoni che escono dalla mia bocca, se visti come parte della fisica, sono emissioni d'aria acustiche piuttosto banali. Ma esse hanno delle caratteristiche notevoli. Innanzitutto, noi pensiamo che esse siano delle frasi in italiano e che i loro enunciati siano degli atti linguistici. Se noi pensiamo ad esse come a frasi e atti linguistici, e se noi tutti cooperiamo nell'usarle, nell'interpretarle, nel considerarle, nel rispondervi e nel trattarle, in generale, come frasi e come atti linguistici allora esse sono realmente ciò che noi usiamo, consideriamo, trattiamo e interpretiamo. (Sono stato molto breve su questo punto. Ma non vorrei si pensasse che questi fenomeni siano dei fenomeni semplici). In questi casi noi creiamo una realtà istituzionale trattando la nuda realtà come avente un certo *status*. Le entità in questione – il denaro, la proprietà, il governo, il matrimonio, le università e gli atti linguistici – hanno tutte un livello di descrizione analogo a quello di semplici fenomeni fisici come le montagne e i cumuli di neve. Ma attraverso l'intenzionalità collettiva noi

⁴ *The Construction of Social Reality*, New York, The Free Press, 1995, tr. it. *La costruzione della realtà sociale*, Edizioni di Comunità, Milano 1996.

imponiamo degli *status* su di esse, e con questi *status* imponiamo delle funzioni che esse non potrebbero compiere senza tale imposizione.

Il prossimo passo è osservare come nella creazione di questi fenomeni istituzionali noi possiamo anche creare delle ragioni per l'azione. Io ho una ragione per conservare e mantenere dei pezzi di carta poco interessanti nel mio portafoglio perché so che essi sono qualcosa più di semplici pezzi di carta. Essi sono pezzi di carta di valuta corrente degli Stati Uniti. Ovvero, data la struttura istituzionale, vi sono insiemi di ragioni per le azioni che non potrebbero esistere senza la struttura istituzionale. Così, l'enunciato "sembra il caso che" può creare un insieme di ragioni per l'azione, poiché ciò che sembra il caso che (se compreso nel modo giusto), è proprio il caso in cui è coinvolta la realtà istituzionale. Se io mi faccio prestare del denaro da qualcuno, o ordino una birra in un bar, o mi sposo, o mi iscrivo come socio ad un club, io uso le strutture istituzionali per creare delle ragioni per l'azione e le ragioni esistono all'interno delle strutture istituzionali.

Ma tutto questo non risponde ancora alla nostra domanda cruciale, e cioè: come possiamo usare tali strutture per creare delle ragioni indipendenti dal desiderio? Ho delle ragioni molto buone per volere del denaro, ma esse sono tutte dipendenti dal desiderio, poiché derivano dal desiderio che provo nei confronti delle cose che posso acquistare con il denaro. Ma che dire dei miei obblighi a pagare del denaro? O del pagare i miei debiti ad altre persone? O del mantenere le mie promesse di dare del denaro in determinate occasioni? Se un gruppo di persone creasse un'istituzione la cui sola funzione fosse quella per cui io dovrei dare loro del denaro, io non avrei, nemmeno allora, nessun obbligo di dare loro del denaro, poiché, sebbene essi possano aver creato ciò che pensano sia una ragione, questa non è ancora una ragione per me. In che modo posso dunque utilizzare la realtà istituzionale per creare delle ragioni indipendenti dal desiderio per me?

È a questo punto che dobbiamo introdurre le caratterizzazioni della libertà e il punto di vista della prima persona. La nostra domanda ora è come posso creare una ragione per me stesso, una ragione che mi vincolerà nel futuro, anche se a quel tempo potrei non avere nessun desiderio di fare la cosa per la quale avevo creato una ragione per farla. Io penso sia impossibile dare una risposta a questa domanda se si guarda ai fenomeni dal punto di vista della terza persona. Da un punto di vista della terza persona, qualcuno emette una serie di suoni dalla bocca. Egli dice "prometto che ti sveglierò alle 6:00". Come potrebbe mai questo fatto creare una ragione che vincolerà la sua volontà? Il solo modo per rispondere a questa

domanda è osservare, dal punto di vista della prima persona, che cosa sta accadendo, che cosa sto cercando di fare, qual è la mia intenzione quando emetto quei suoni dalla mia bocca. E una volta che abbiamo osservato la questione dal punto di vista della prima persona, noi possiamo, credo, vedere la soluzione per il nostro rompicapo. Quando dico “ti prometto che ti sveglierò alle 6:00” io vedo me stesso come un soggetto che crea liberamente un tipo speciale di ragione indipendente dal desiderio, un obbligo, per me, di svegliarti alle 6:00. Questa è l’intera questione per quanto riguarda il promettere. Infatti questo è ciò che una promessa è. Essa è la creazione intenzionale di un certo tipo di obbligo – e tali obblighi sono per definizione indipendenti dai successivi desideri dell’agente. Ma quanto ho affermato finora è soltanto che io ho emesso dei suoni con certe intenzioni e che poiché ho quelle intenzioni, mi sembra che le cose stiano proprio in questo modo. Ma come passiamo da “sembra il caso” a “è il caso”? Per rispondere a questa domanda dobbiamo tornare a ciò che ho detto a proposito delle strutture istituzionali. È una caratteristica di queste strutture il fatto che *sembra* è precedente ad *è*. Se mi sembra di creare una promessa, poiché quella era la mia intenzione mentre compivo quell’atto e a voi sembra di aver ricevuto una promessa, e tutte le condizioni ad essa connesse (che non elencherò qui ma che ho elencato in dettaglio altrove),⁵ e se tutte le altre condizioni per la possibilità di creare una promessa sono presenti, allora io ho creato una promessa. Io ho creato intenzionalmente una nuova entità che mi vincola per il futuro; è per me una ragione indipendente dal desiderio, poiché io l’ho creata liberamente e intenzionalmente in quanto tale.

L’abilità di vincolare la volontà ora crea una ragione per l’atto futuro solamente perché essa è una manifestazione della mia libertà in questo momento. Ho detto in precedenza che questo mostra perché lo schiavo non ha nessuna ragione per ubbidire al padrone, se non delle ragioni dipendenti dal desiderio, sebbene sia colui che fa una promessa che lo schiavo agiscano all’interno di strutture istituzionali. Le sole ragioni che lo schiavo possiede sono quelle prudenziali. Lo schiavo non ha mai esercitato alcuna libertà nel creare una ragione per se stesso per agire. Per osservare in che modo all’interno della struttura istituzionale un’agente possa creare delle ragioni esterne per l’agire, è essenziale notare che all’interno della struttura istituzionale l’agente ha la possibilità di creare liberamente delle ragioni per se stesso. Non può esservi nessun dubbio che questa sia una ragione per lui poiché egli l’ha liberamente e volontariamente creata come una ragione per

⁵ *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, Cambridge University Press, 1969, cap.3, tr.it.

se stesso. Questo non significa naturalmente dire che questa è una ragione che sovrasterà tutte le altre ragioni. Al contrario, noi sappiamo che in ogni situazione della vita reale deve esserci probabilmente un gran numero di ragioni in competizione tra loro, a favore o contrarie all'effettuazione di una certa azione. Quando giunge il momento per decidere di fare o di non fare qualcosa, l'agente deve ancora soppesare la sua promessa in relazione a tutti gli altri tipi di ragioni in competizione.

Finora abbiamo considerato quattro caratteristiche: il tempo, le strutture istituzionali, il punto di vista della prima persona e la libertà. Ora ci rivolgeremo alla quinta: la razionalità. L'abilità di agire razionalmente è data da un insieme generale di capacità che comprendono cose come l'abilità di riconoscere e operare con coerenza, l'inferenza, il riconoscimento dell'evidenza e un gran numero di altre capacità. Le caratteristiche della razionalità che sono importanti per la presente discussione implicano la capacità di operare in vari modi con le ragioni per l'azione. Vorrei che a questo punto della trattazione questo risultasse vago perché il chiarificarlo sarà il nostro prossimo, essenziale, compito.

Supponiamo che io abbia agito liberamente con l'intenzione di creare una ragione indipendente dal desiderio per me, supponiamo che io abbia soddisfatto tutte le condizioni (del promettere, dell'ordinare una birra o qualsiasi altra cosa), di modo che io sia effettivamente riuscito a creare quella ragione. Quando arriva il momento, di che cosa ho bisogno allora per riconoscere che c'è una tale ragione? Assumendo che io conosca tutti i fatti, la razionalità riconoscitiva è sufficiente per venire a conoscenza del fatto che la creazione precedente di una ragione è ora vincolante. La cosa importante è che non è necessario possedere un principio morale ulteriore riguardo al promettere o al bere la birra per capire che la ragione che si è creata in passato come ragione vincolante per il momento attuale è proprio una ragione vincolante nel momento attuale. È una pura incoerenza logica ammettere tutti i fatti, riguardanti la creazione e la continuazione dell'obbligo, e poi negare di avere una ragione per agire.

VI. Sommario e conclusioni

In questo capitolo mi sono occupato di mostrare in che modo gli esseri umani possono creare ed essere motivati ad agire sulla base di ragioni per l'azione indipendenti dal

desiderio. Quali fatti corrispondono alla tesi secondo la quale l'agente ha creato una tale ragione, e quali fatti corrispondono alla tesi che tale ragione sia una forma razionale di motivazione all'azione? Ho provato a discutere tali questioni a tre livelli. Il primo livello, e il più basilare, è quello della struttura atomica dell'intenzionalità fondamentale tramite la quale un agente può impegnare se stesso imponendo condizioni di soddisfazione su altre condizioni di soddisfazione. Il secondo livello è il livello della "fenomenologia" all'interno del quale noi discutiamo di ciò che appare all'agente. Ciò che appare all'agente è che egli assume degli impegni attraverso il libero e intenzionale esercizio della sua volontà, in modo tale da legare la sua volontà al futuro così che in futuro egli avrà una ragione per un'azione che sarà indipendente dal desiderio di compiere l'azione o meno. E il terzo livello è quello della società in generale – quali sono le funzioni sociali di tali sistemi di ragioni per l'azione indipendenti dal desiderio?

I fatti basilari che corrispondono alla tesi secondo cui gli esseri umani possono creare ed essere motivati ad agire sulla base di ragioni indipendenti dal desiderio sono i seguenti:

1. Deve esistere una struttura sufficiente per la creazione di tali fatti istituzionali. Queste strutture sono inevitabilmente linguistiche ma possono anche implicare altre istituzioni. Tali strutture ci permettono di acquistare una casa, di ordinare una birra, di iscriversi all'università, ecc.
2. All'interno di queste strutture, se l'agente agisce con intenzioni appropriate, questo è sufficiente per la creazione di ragioni indipendenti dal desiderio. Specificamente, se l'agente agisce con l'intenzione che la sua azione dovrebbe creare una tale ragione, allora se le circostanze sono quelle appropriate, egli ha creato quella ragione. L'intenzione cruciale è l'intenzione che essa sia una ragione. La ragione non deriva dall'istituzione; l'istituzione fornisce solamente il veicolo per la creazione di tali ragioni.
3. La forma logica dell'intenzionalità nella creazione di tali ragioni è inevitabilmente l'imposizione di condizioni di soddisfazione su altre condizioni di soddisfazione. Il caso più lampante, per così dire, della creazione di una ragione per un'azione indipendente dal desiderio è quello della promessa. Il promettere è comunque peculiare tra gli atti linguistici in quanto l'autore della promessa è il soggetto del contenuto proposizionale ed ha una componente autoreferenziale imposta sulle

condizioni di soddisfazione. Le condizioni di soddisfazione della promessa non sono solamente che il parlante faccia qualcosa, ma che egli faccia quel qualcosa perché ha promesso di farlo. C'è, quindi, una componente autoreferenziale nel promettere, e questa componente autoreferenziale non esiste in certi altri tipi di atti linguistici. Per esempio, non esiste nelle asserzioni.

4. Una volta che l'obbligo è creato, è un requisito della razionalità riconoscitiva che l'agente debba riconoscerlo come vincolante il suo comportamento seguente. L'obbligo ha la struttura delle ragioni per l'azione. C'è un'entità fattiva con direzione di adattamento verso l'alto e l'agente come soggetto.
5. Una volta che sia stata creata una valida ragione per l'azione indipendente dal desiderio, quella ragione può motivare un desiderio di compiere l'azione, proprio come il riconoscimento di qualsiasi altra ragione può motivare un desiderio di compiere l'azione. Riconoscere una ragione valida per fare qualcosa è già riconoscere una ragione valida per volerla fare.

Appendice al capitolo 6: ragioni interne ed esterne

Ho posto un'obiezione alla tesi di Bernard Williams secondo la quale non esistono cose come le ragioni esterne, poiché tutte le ragioni per un agente devono essere interne al suo insieme motivazionale. Senza dubbio ci sono varie obiezioni che si potrebbero avanzare a questa tesi ma la caratteristica, e la forza, principale della mia obiezione è stata che possono esserci dei fatti esterni all'insieme motivazionale dell'agente tali che la razionalità richieda che l'agente riconosca questi fatti come ragioni per l'azione, anche se non c'è nulla nel suo insieme motivazionale proprio in quel momento che lo disponga a riconoscerli in quanto ragioni. I due tipi di fatti su cui mi sono concentrato sono fatti che riguardano la prudenza a lungo termine e l'esistenza di ragioni indipendenti dal desiderio come gli obblighi assunti dall'agente.

C'è un'ultima caratteristica della dottrina dell'internalismo che merita una speciale menzione. Vi sono interpretazioni dell'internalismo secondo le quali la tesi per cui non vi sarebbero delle ragioni esterne risulta tautologicamente vera, e non vorrei che si pensasse che io sia in disaccordo con tali interpretazioni. Il problema è che le vere descrizioni

tautologiche possono essere facilmente interpretate come descrizioni sostanziali, le quali sono false. (Non sto suggerendo che Williams stesso abbia fatto questa confusione). E in questa appendice ho intenzione, il tutto troppo brevemente, di esporre alcune descrizioni tautologiche e di porle in contrasto con le descrizioni sostanziali.

L'argomento di fondo dell'internalismo è che se un agente non ha delle ragioni interne egli non ha nulla *per cui* ragionare. Una ragione esterna è, per definizione, una ragione che è esterna all'agente, e, di conseguenza, essa è una ragione della quale egli non potrebbe ragionare. Un corollario a questo argomento, e, in un certo senso il modo più potente per esporre l'argomento stesso, è che noi non possiamo spiegare l'azione di un agente nei termini delle sue ragioni a meno che esse non siano delle ragioni interne, poiché solamente una ragione interna può effettivamente *motivare* l'agente ad agire. Così ci sono due argomenti strettamente correlati per l'internalismo: uno che riguarda il processo del ragionare e uno che riguarda la motivazione. Ognuno di questi ammette una formulazione tautologica e, naturalmente, io non posso essere in disaccordo con le formulazioni tautologiche.

Tautologia A, il ragionare: per poter ragionare mentalmente sulla base di una ragione, un agente deve avere una ragione in mente da cui sviluppare il ragionamento.

La versione tautologica di una tesi motivazionale è la seguente:

Tautologia B, motivazione: per essere motivato da una ragione mentale, un agente deve avere una ragione mentale che lo motivi.

Entrambe queste tautologie ammettono una riformulazione sostanziale che non mi sembra tautologica, ma falsa. La riformulazione sostanziale incorpora il disaccordo tra gli internalisti e gli externalisti, per quanto riguarda la razionalità.

Tesi sostanziale A: affinché un qualsiasi fatto o entità fattiva *R* sia una ragione per l'agente *X*, *R* deve già essere parte, o essere rappresentato, dell'insieme motivazionale *S* di *X*.

E la descrizione non tautologica di B è:

Tesi sostanziale B: tutte le motivazioni razionali sono dei desideri in senso generale, nel modo in cui Williams descrive S.

Le descrizioni sostanziali dell'internalismo sono immediatamente soggette a dei controesempi. La tesi A implica l'immediata conseguenza che i fatti che riguardano le ragioni per l'azione indipendenti dal desiderio di un agente, come i fatti che riguardano i suoi interessi prudenziali e i fatti che riguardano le sue contrattazioni e i suoi obblighi, non possono essere delle ragioni per l'azione, anche nei casi in cui l'agente è consapevole di questi fatti, a meno che l'agente sia disposto nel suo insieme motivazionale ad agire sulla base di questi fatti. La tesi B ha l'immediata conseguenza che in qualsiasi momento della vita di un agente, e per ogni atto di tipo *T* - a meno che l'agente proprio in quel momento abbia un qualche desiderio, inteso in senso generale, che sia il compiere un atto di tipo *T*, o il desiderare qualcosa manifestando un chiaro percorso deliberativo che va da quel desiderio al compimento di un atto di tipo *T* come mezzo per soddisfare il desiderio - l'agente non ha alcuna ragione per compiere un atto di tipo *T*. Abbiamo presentato una serie di casi in cui questo è falso, in cui l'agente ha una ragione per compiere un'azione anche quando queste condizioni non vengano soddisfatte.

La disputa tra internalisti ed externalisti riguarda dunque l'esistenza di ragioni per l'azione indipendenti dal desiderio. La domanda è la seguente: esistono ragioni tali per cui la sola razionalità può essere sufficiente perché l'agente le riconosca in quanto motivazioni, indipendentemente dal fatto che esse si riferiscano a qualcosa che sta nell'insieme motivazionale dell'agente? Secondo l'internalista tutte le ragioni per l'azione devono essere fondate sui desideri, intesi in senso generale. Secondo l'externalista ci sono alcune ragioni per l'azione che possono esse stesse costituire il fondamento del desiderio di fare qualcosa, ma queste non sono né dei desideri né sono basate sui desideri. Per esempio, io posso avere il desiderio di mantenere la mia promessa perché la riconosco come un obbligo, senza per questo pensare che la sola ragione per cui la voglio mantenere sia il fatto che, precedentemente, io avevo il desiderio di mantenere tutte le mie promesse.

Williams sembra talvolta sostenere che il riconoscimento di un obbligo è già una ragione interna per l'azione. Ma questa tesi è ambigua. Dire che A sa di avere un obbligo consente almeno due distinte possibilità.

1. A sa di avere un obbligo, che egli riconosce come una ragione valida per agire e quindi come una ragione per voler agire.
2. A sa di avere un obbligo, ma questo non gli importa minimamente. Niente nel suo insieme motivazionale lo rende incline ad agire sulla base di esso.

Ora la disputa tra internalisti ed externalisti si sviluppa proprio su questo punto: per gli internalisti, vi sono in entrambi i casi delle ragioni per l'azione. Infatti in entrambi i casi vi sono delle ragioni per l'azione indipendenti dal desiderio. Per l'internalista, solamente nel caso (1) c'è una ragione per l'azione. Inoltre, secondo l'externalista, il caso (1) è descritto in maniera fuorviante dall'internalismo. L'internalista pensa che il riconoscimento di un obbligo vincolante come una ragione valida sia già un desiderio per l'azione. L'externalista pensa a questo come al fondamento di un desiderio, che è esso stesso una ragione per l'azione indipendente dal desiderio.

In tali casi mi sembra che il difensore del punto di vista internalista potrebbe argomentare che la ragione esterna può funzionare soltanto se l'agente ha la capacità di riconoscerla come un obbligo vincolante. E questo conduce alla terza descrizione tautologica dell'internalismo:

Tautologia C: nell'esercizio delle sue capacità disposizionali interne, affinché un agente riconosca una ragione esterna come una ragione, l'agente deve avere la capacità interna di riconoscerla come una ragione.

Ma questo può essere facilmente reinterpretato in una versione sostanziale non tautologica, la quale è falsa:

Descrizione sostanziale C: affinché un qualsiasi fatto esterno possa essere una ragione per l'agente, l'agente deve avere una disposizione interna a riconoscerlo come ragione.

È facile notare come si possa confondere il sostanziale con il tautologico, ma essi sono piuttosto distinti. Il tautologico dice semplicemente che per esercitare una certa capacità l'agente deve possedere la capacità in questione. La descrizione sostanziale dice

che nulla è una ragione valida a meno che l'agente sia disposto a riconoscerla come tale e questo, come ho sostenuto, è errato. È parte del concetto di razionalità che possano esservi delle ragioni indipendenti dal desiderio, ragioni che vincolano un agente razionale, indipendentemente dal desiderio e dalle disposizioni che si trovano nel suo insieme motivazionale.

P 12

7

LA DEBOLEZZA DELLA VOLONTÀ

Alcune volte, per la verità piuttosto spesso, può accadere che si affronti un processo di deliberazione, che si prenda una decisione ponderata, con cui si forma un'intenzione ferma e incondizionata a fare qualcosa e, quando arriva il momento, a causa della debolezza della volontà, non la si faccia. Ora, se la relazione tra deliberazione e intenzione fosse sia causale che razionale o logica, cioè, se i processi razionali causassero le intenzioni e se, a loro volta, le intenzioni causassero le azioni tramite la causazione intenzionale, allora come potrebbero mai esistere dei casi genuini di debolezza della volontà? Come può essere che un agente formi un'intenzione fortissima, incondizionata di fare qualcosa, che nulla glielo impedisca, e che tuttavia egli non la faccia? Sorprendentemente, molti filosofi pensano che una cosa del genere sia impossibile ed hanno avanzato argomenti ingegnosi per dimostrare che è impossibile, e che i casi apparenti di debolezza della volontà sono, in realtà, dei casi che riguardano qualcos'altro. Ma, ahimè, questo non solo è possibile, ma anche piuttosto comune. Ecco l'esempio di un caso comunissimo: uno studente si forma un'intenzione ferma e incondizionata di lavorare al suo compito per l'esame martedì sera. Non vi è nulla che gli impedisca di lavorare, ma quando arriva la mezzanotte si scopre che ha trascorso la sera guardando la televisione e bevendo della birra. Tali casi, come ogni insegnante può attestare, non sono affatto insoliti. Dovremmo in effetti insistere sul fatto che questa è una condizione di normalità in qualsiasi spiegazione della debolezza della volontà, di quella che i Greci chiamavano *akrasia* (incontinenza), e che fa sì che l'*akrasia* sia molto comune nella vita reale e non implichi alcun errore logico. Nei primi capitoli abbiamo scoperto un divario tra le intenzioni e le azioni, e questo divario (*gap*) ci fornirà la spiegazione della debolezza della volontà.

Ebbene, come è possibile l'*akrasia*? Proviamo ad aggirare la domanda e a chiederci perché si dovrebbe dubitare della sua possibilità e, esserne addirittura sconcertati, dal momento che nella vita reale essa è così comune? Penso che l'errore fondamentale, ed è un errore che ha una lunga storia nel pensiero filosofico, sia quello di porre in modo errato la relazione tra gli antecedenti e il compimento di un'azione. C'è una lunga tradizione in filosofia secondo la quale, nei casi dell'azione razionale, se gli antecedenti psicologici di un atto sono tutti in un certo ordine, e cioè sono il tipo giusto di desideri, intenzioni, giudizi di valore, ecc., allora l'atto deve necessariamente seguire. Secondo alcuni autori si tratta perfino di una verità analitica. Una tipica espressione dell'idea di necessità causale si trova in J.S. Mill:

... le volizioni in effetti seguono determinati antecedenti morali con la medesima uniformità e (quando abbiamo una conoscenza sufficiente delle circostanze) con la stessa certezza con cui gli effetti fisici seguono le loro cause fisiche. Questi antecedenti morali sono i desideri, le avversioni, le abitudini e le disposizioni, e, combinati con le circostanze esterne, essi si potrebbero definire gli incentivi interni all'azione... una volizione è un effetto morale che segue delle corrispondenti cause morali così certe e invariabili quanto lo sono gli effetti fisici che seguono le cause fisiche.¹

Penso sia chiaro che chiunque sostenga una tale tesi considererà la debolezza della volontà un problema, perché se le cause sono del tipo appropriato, allora l'azione dovrebbe seguire per necessità causale. C'è una tradizione nella filosofia analitica del ventesimo secolo secondo la quale dei casi puri di debolezza della volontà in realtà non accadono mai, ed è impossibile che essi possano accadere. Secondo R.M. Hare², se l'agente agisce contrariamente alla sua convinzione morale professata, significa che egli non possedeva realmente la convinzione morale che sosteneva di avere. Secondo Donald Davidson³, se l'agente agisce contrariamente alle sue intenzioni, allora non aveva realmente un'intenzione incondizionata di compiere l'azione. Sia Hare che Davidson sostengono alcune varianti dell'idea di base secondo la quale colui che esprime un certo tipo di giudizio valutativo in favore del fare una certa cosa deve allora necessariamente fare quella cosa (a meno che, naturalmente, non ne sia impedito). Di conseguenza, secondo questa prospettiva, se l'azione non viene compiuta, significa semplicemente che non era presente un giudizio valutativo del

¹ J.S. Mill, *The Examination of Sir William Hamilton's Philosophy*, citato in Timoty O'Connor (ed.), *Agent, Causes and Events: Essays on Indeterminism and Free Will*, Oxford, Oxford University Press, 1995, p. 76.

² R.M. Hare, *The Language of Morals*, Oxford, Oxford University Press, 1952, tr. it. *Il linguaggio della morale*, Ubaldini, Roma 1968.

³ "How Is Weakness of the Will Possible?" in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, tr. it. "La debolezza della volontà", in *Azioni ed eventi*, il Mulino, Bologna 1992, pp. 63-88.

giusto tipo. Dalla spiegazione di Davidson si evince che il giudizio era solamente un giudizio di valore *prima facie* o condizionale. Dalla spiegazione di Hare si evince che la valutazione in questione non avrebbe potuto essere una valutazione morale.

Il modello generale, in tutti questi casi, è supporre che se gli antecedenti dell'azione sono strutturati razionalmente in un certo modo, allora l'azione seguirà per necessità causale. Davidson sostiene dunque i due seguenti principi:

(P1) Se un agente vuole fare x più intensamente di quanto voglia fare y ed egli si ritiene libero di fare sia x che y , allora farà intenzionalmente x se fa intenzionalmente o y o x . (Ibid., p. 65)

e

(P2) Se un agente giudica che sarebbe meglio fare x piuttosto che fare y allora vorrà fare x più di quanto voglia fare y . (Ibid.)

Presi insieme questi principi implicano che un agente che giudica che sarebbe meglio fare x piuttosto che y farà intenzionalmente x se farà intenzionalmente o x o y . Questi due principi sembrano essere inconsistenti con il principio secondo il quale ci sono atti di volontà debole, e che Davidson afferma in questo modo:

(P3) Vi sono azioni incontinenti. (Ibid.)

Ovvero, ci sono azioni in cui l'agente giudica che sarebbe meglio fare x piuttosto che y , ritiene se stesso libero di fare x e tuttavia, intenzionalmente, fa y invece che x .

La soluzione di Davidson dell'apparente paradosso è dire che i casi in cui l'agente apparentemente agisce in modo contrario al suo migliore giudizio facendo y piuttosto che x sono in realtà casi in cui l'agente non aveva formulato un giudizio *incondizionato* secondo il quale il miglior corso delle azioni sarebbe stato fare x . La tesi di Hare è leggermente più complessa, ma esprime la stessa idea di fondo. La sua idea è che se noi accettiamo un imperativo, o un comando, allora segue per necessità causale che il nostro accettare quell'imperativo porterà al compimento dell'azione e, secondo la sua teoria, accettare un giudizio morale significa accettare un imperativo. Hare scrive:

“Proponiamo che, per stabilire se uno usa il giudizio ‘Devo fare X’ come un giudizio valutativo oppure no, ci si ponga la seguente domanda: “Ammette o non ammette egli che se assente a quel giudizio deve necessariamente assentire anche al comando ‘Che io faccia X’?”⁴.

Egli scrive inoltre che “è tautologico dire che non possiamo dare il nostro sincero assenso a un comando rivoltoci in seconda persona e nello stesso tempo non eseguirlo, se ora è il momento di eseguirlo e ci è (fisicamente e psicologicamente) possibile eseguirlo”.⁵

Entrambi gli autori sostengono la tesi secondo la quale gli appropriati antecedenti causali dell’azione devono causare l’azione, quindi i casi apparenti di *akrasia* sono in realtà casi in cui c’era qualcosa di sbagliato nelle cause dell’azione nella forma di stati psicologici antecedenti.

Questi autori negano in realtà l’esistenza del divario (*gap*), e questo è il motivo per cui il problema della debolezza della volontà emerge per loro in forma così rigida e per cui essi sono costretti a negare, sia implicitamente che esplicitamente, che esistano davvero cose come i casi di *akrasia* in senso stretto. Così, la disputa profonda che c’è tra il mio pensiero e la tradizione filosofica è una disputa che riguarda il divario. Il modello classico nega l’esistenza del divario. Io, al contrario, credo che il divario sia un fatto evidente della nostra vita cosciente. Ho fornito degli argomenti a favore dell’esistenza del divario nei primi capitoli e non intendo riproporli qui. In questo capitolo voglio adottare un approccio diverso. Io considero l’approccio di Hare-Davidson all’*akrasia* una specie di *reductio ad absurdum* di questa caratteristica del modello classico. Secondo la mia prospettiva, nel caso delle azioni libere, non importa quale tipo di antecedenti abbia l’azione – giudizi morali, giudizi di valore incondizionati, intenzioni ferme e incondizionate o qualsiasi cosa vogliate – la debolezza della volontà è sempre possibile. Così, se si arriva alla conclusione che non essa è possibile, si deve aver commesso un errore e quindi bisogna retrocedere e andare a rivedere le premesse che hanno portato all’errore. In questo caso la premessa falsa è la negazione del divario. La spiegazione di Davidson è più recente, quindi focalizzerò gran parte della mia attenzione su di essa.

⁴ R.M. Hare, *The Language of Morals*, Oxford, Oxford University Press, 1952, pp. 168-169, tr. it. *Il linguaggio della morale*, Ubaldini, Roma 1968, p. 153.

⁵ Ibid., p. 30.

Che cosa sostiene, esattamente, la tesi secondo la quale ci sono delle azioni debolmente volute? Cioè, abbiamo bisogno di enunciare la tesi in un modo che renda chiaro se essa è veramente inconsistente con (P1) e con (P2). Davidson la esprime in questa forma:

(P3) Ci sono delle azioni incontinenti.

Ma che cos'è un'azione "incontinente"? Secondo un'interpretazione naturale, mi sembra che la tesi affermi che ci sono degli atti per i quali l'agente giudica incondizionatamente che sarebbe meglio fare *x* piuttosto che *y*, ritiene di essere in grado di fare entrambi, ma tuttavia, intenzionalmente, fa *y* piuttosto che *x*. Questa tesi è genuinamente inconsistente in relazione a (P1) e (P2), ed è una tesi che io ritengo essere vera. Davidson nega che essa sia vera e sostiene che nei casi che sembrano essere "incontinenti" ciò che effettivamente accade è il fatto che l'agente non aveva giudicato incondizionatamente che sarebbe stato meglio fare *x* piuttosto che *y*, ma, piuttosto, aveva soltanto emesso un giudizio condizionale, o *prima facie*, il cui contenuto era che *x* era migliore di *y*. Egli aveva giudicato che *x* fosse migliore di *y* "tutto considerato", dove, secondo Davidson, "tutto considerato" non significa letteralmente "considerate tutte le possibilità", ma significa semplicemente qualcosa del tipo "relativamente ad un certo insieme di considerazioni che l'agente aveva in mente".

La prima cosa da notare riguardo alla tesi di Davidson è che non viene fornito nessun argomento indipendente per sostenere che l'agente con una volontà debole non può emettere un giudizio di valore incondizionato a favore dell'esecuzione di una qualsiasi altra azione diversa da quella che compie. Cioè, non viene fornita nessuna ragione indipendente per motivare la tesi, non viene esaminato nessun caso per dimostrare che era stato espresso solamente un giudizio condizionale. Piuttosto, la nozione di valutazioni *prima facie* e condizionate viene introdotta come un modo per superare l'apparente inconsistenza tra (P1), (P2) e (P3). Se nel caso delle azioni con volontà debole l'agente non ha espresso un giudizio incondizionato in favore del fare un'azione che poi non ha fatto, ma solamente un giudizio *prima facie*, "tutto considerato", allora l'inconsistenza viene rimossa. Per ora (P3) si legge come segue:

(P3*) A volte un agente esprime un giudizio condizionale, *prima facie*, secondo il quale sarebbe meglio fare *x* piuttosto che *y*, ritiene di essere in grado di fare entrambe le cose, e poi, intenzionalmente, fa *y*.

Se intese in questo modo (P1), (P2) e (P3) sono consistenti.

Qual è allora lo statuto logico della soluzione? La tesi è questa: tutte le azioni con volontà debole sono precedute da dei giudizi di valore condizionali (o da intenzioni condizionali, che Davidson ritiene siano la stessa cosa). Posta in tale modo, questa sembra un'ipotesi empirica: c'è il cento per cento di correlazione tra l'esperienza della debolezza della volontà e l'espressione di giudizi di valore condizionali piuttosto che incondizionati. Ma se si suppone che questa sia un'ipotesi empirica, si tratta di una tesi incredibilmente ambiziosa, costruita sulla base di una minima, quasi nessuna, evidenza empirica.

E perfino al di là del fatto che non viene fornito nessun argomento per sostenere che l'agente con volontà debole non ha espresso un giudizio incondizionato, c'è ancora un altro, e più grave, problema. Il problema è questo: non importa quale sia la forma del giudizio, un agente può sempre soffrire per la debolezza della volontà. Un agente può dire “penso incondizionatamente che *x* sia meglio di *y*”, e, ciononostante, fare *y* invece che *x*. Il solo modo in cui posso uscirne è di rendere circolare l'argomento, far sì che il criterio per verificare se una persona aveva un giudizio incondizionato o meno sia il verificare se egli ha compiuto l'azione intenzionalmente oppure no. Il circolo è questo: la tesi è che tutte le azioni con volontà debole sono precedute da intenzioni condizionali, invece che da intenzioni incondizionate. L'argomento a sostegno della tesi è che le azioni avevano volontà debole, e che perciò *dovevano* essere state precedute da un'intenzione condizionale, piuttosto che da un'intenzione incondizionata, poiché se fossero state precedute da un'intenzione incondizionata, l'azione avrebbe dovuto avere luogo. Credo che questo circolo sia implicito nell'articolo di Davidson. Secondo Davidson, un agente fa qualcosa intenzionalmente se e solo se fa proprio un fortissimo giudizio di valore incondizionato in favore del fare quella cosa. Da questa concezione segue banalmente il fatto che nei casi in cui l'agente dice che *x* è meglio di *y*, ma tuttavia compie intenzionalmente *y* invece che *x*, il giudizio non poteva essere incondizionato. Ma questo ci fa passare dalla padella alla brace poiché è ovviamente falso, per qualsiasi senso ordinario dell'esprimere giudizi di valore fortissimi, il fatto che sia impossibile per qualcuno esprimere un tale giudizio e poi non fare

la cosa che egli giudica meglio fare. Infatti è esattamente questo il problema della debolezza della volontà. Spesso si emette un giudizio incondizionato fortissimo e poi non si fa la cosa che si giudica come la migliore cosa da fare. Davidson, semplicemente, risolve il problema della debolezza della volontà con un decreto quando dichiara che in tali casi l'agente non riesce ad esprimere un giudizio incondizionato fortissimo.

La mia diagnosi di quanto accade è la seguente: quella che sembra una tesi empirica – tutti i casi di debolezza della volontà sono casi di giudizi di valore condizionali, o condizionati – non è di fatto una tesi empirica. Piuttosto, Davidson assume che (P1)-(P3) siano vere e che (P1) e (P2) non siano problematiche e, quindi, ci deve essere un'interpretazione di (P3) in cui essa è inconsistente con (P1) e (P2). La tesi circa i giudizi di valore condizionali corrisponde a tale interpretazione.

Ma questa soluzione ha delle conseguenze assurde, che ora esporrò. Consideriamo quei casi di debolezza della volontà che normalmente si presentano nella vita reale. Supponiamo che io decida, dopo aver considerato tutti i fatti che conosco riguardanti il problema, che sia meglio per me non bere del vino stasera a cena, perché, supponiamo, io voglio lavorare al problema della debolezza della volontà dopo cena. Ma supponiamo che, come accade, io beva del vino a cena. Una volta servito il vino, esso sembrava così attraente che, in un momento di debolezza, non ho resistito alla tentazione e l'ho bevuto. Secondo la spiegazione di Davidson, questo è l'elenco completo dei miei stati intenzionali che sono coinvolti in questo caso:

1. Ho espresso un giudizio condizionale: tutto considerato, è meglio che io non beva del vino.
2. ho emesso un giudizio incondizionato: è meglio bere del vino.

E, verificandosi quel caso, ho bevuto del vino.

Che cosa c'è di sbagliato in questa spiegazione? È semplicemente falso dire che devo avere espresso un giudizio di valore incondizionato il cui contenuto è che è meglio bere del vino. Ho semplicemente bevuto del vino. Questo è ciò che ha reso la mia azione un caso di debolezza della volontà. Ho bevuto del vino a dispetto del mio giudizio *incondizionato* che è meglio non bere del vino. Così la tesi falsa secondo cui la mia intenzione di fare la cosa giusta non sarebbe stata incondizionata, ma sarebbe stata soltanto condizionata o *prima*

facie, si accorda con un'altra tesi falsa secondo la quale quando ho fatto la cosa sbagliata dovevo avere espresso un giudizio incondizionato il cui contenuto era che quella, in quel momento, era la cosa giusta da fare. Entrambe le tesi sono sbagliate. Io posso esprimere un giudizio di valore incondizionato e tuttavia fare qualcosa di contrario ad esso, con un atto di volontà debole e tale atto di volontà debole non ha bisogno di alcun giudizio che abbia contenuto che quella era la cosa giusta da fare.

Il problema della debolezza della volontà *non* riguarda la possibilità di riconciliare due *giudizi* apparentemente inconsistenti. Il problema è piuttosto il seguente: come è possibile che avendo espresso un solo giudizio io possa in seguito agire contrariamente rispetto a quello stesso giudizio? E la risposta è che io non devo esprimere un altro giudizio per agire; io semplicemente agisco. Cioè, in questo tipo di casi, io ho un'intenzione-in-azione senza nessuna intenzione precedente e senza nessuna deliberazione precedente.

Ciò che l'intera discussione mostra è che la congiunzione di (P1) e (P2) è falsa. Non si dà sempre il caso che ogni cosa che si giudichi come la migliore cosa da fare, sia ciò che realmente si vuole fare, e non si dà nemmeno il caso che quando si è presa una decisione e davvero si vuole fare qualcosa, la si faccia perciò necessariamente. Ci sono molte cose che io considero le migliori cose da fare, e che considero le cose che davvero voglio fare, ma che di fatto non faccio, anche se dispongo sia della capacità che dell'opportunità di farle.

La frase chiave, io credo, nell'articolo di Davidson, è la seguente: "Sia *r* la ragione di qualcuno per giudicare che *p*; allora il suo giudicare che *r* deve essere, così mi pare, una causa del suo giudicare che *p*. Ma ecco il punto cruciale: il suo giudicare che *r* può causare il suo giudicare che *p* senza che *r* sia la sua ragione; anzi, l'agente potrebbe anche pensare che *r* sia una ragione per respingere *p*"⁶. Proviamo ad applicare questa spiegazione all'esempio del bere del vino. Io giudico un insieme di ragioni *r*, e quelle ragioni causano il mio giudicare che è meglio bere del vino, *p*. Esse causano dunque il fatto che io ritengo sia meglio bere del vino senza effettivamente essere una ragione per bere del vino. In realtà, nel mio caso, io penso che esse siano una ragione per respingere l'affermazione che è meglio bere del vino.

Penso che questa non sia nemmeno lontanamente plausibile come spiegazione di ciò che è accaduto quando io, in un momento di debolezza della volontà, ho bevuto del vino a dispetto del mio miglior giudizio. Ritengo che una spiegazione decisamente più plausibile,

⁶ "La debolezza della volontà", tr.it. p. 87.

che spiegherò in maggiore in dettaglio più avanti, sia che io ho sostenuto un giudizio incondizionato il cui contenuto era che fosse meglio *non* bere del vino, ma una volta messo a confronto con il vino, l'ho reputato attraente e semplicemente non sono riuscito a resistere alla tentazione.

Come siamo finiti in questo pasticcio? Davidson in compagnia di molti altri filosofi⁷, pensa che nel caso delle azioni motivate razionalmente, vi sia un qualche tipo di connessione causalmente necessaria tra gli antecedenti psicologici di un'azione e il compimento intenzionale dell'azione, o almeno il tentativo intenzionale di compiere l'azione: egli ritiene che l'azione segua dai suoi antecedenti per una sorta di necessità causale. Ma questo è un errore. Questo nega l'esistenza del divario. E una volta negata l'esistenza del divario, si cade in tutti quei problemi che abbiamo esaminato, e in particolare si cade nel problema per cui la debolezza della volontà, strettamente parlando, diventa impossibile.

In risposta a queste tesi secondo le quali gli appropriati antecedenti psicologici portano all'azione in questione grazie ad una necessità causale, chiediamoci: esistono in realtà casi del genere? Esistono casi in cui gli antecedenti psicologici sono causalmente sufficienti a produrre l'azione? Mi sembra piuttosto ovvio che vi siano molti casi di questo tipo, ma sono normalmente casi in cui non c'è alcun libero arbitrio, piuttosto che casi di azioni volontarie di tipo comune. Così, per esempio, un tossicodipendente potrebbe avere benissimo degli antecedenti psicologici per prendere della droga che sono causalmente sufficienti a garantire il suo uso di droga, semplicemente perché non è capace di farne a meno. In questi casi, come abbiamo visto prima, non c'è un divario di tipo familiare. L'azione è determinata causalmente da sufficienti cause psicologiche antecedenti. Oggi, a tale riguardo, disponiamo inoltre di una buona evidenza che tali cause psicologiche abbiano un fondamento di tipo neurobiologico. Quando una persona manifesta la necessità di dover soddisfare la propria dipendenza dalla droga, viene attivato il sistema dopaminico mesolimbico. Questo sistema si estende dall'amigdala e dal cingolato anteriore fino ad entrambi i lobi temporali. La sua attivazione costituisce, almeno secondo alcune teorie attuali, il correlato neurologico del comportamento di dipendenza.

⁷ Per esempio, Peter Van Inwagen, "When Is the Will Free?" in Timothy O'Connor (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, New York, Oxford University Press, 1995.

In casi normali potremmo muovere l'ovvia obiezione che si può esprimere qualsivoglia giudizio di valore e tuttavia non agire sulla base di quel giudizio. Il problema dell'*akrasia*, ripeto, è che se noi tralasciamo i casi di dipendenza, di costrizione, di ossessione, ecc., allora qualsiasi antecedente, posto che esso sia descritto in modo diverso da una petizione di principio che non implichi banalmente l'esecuzione dell'azione, è tale che è sempre possibile per un agente razionale pienamente cosciente avere un antecedente (per esempio, il giudizio morale rilevante, l'intenzione incondizionata, o qualsiasi cosa vogliate) e, tuttavia, non agire in accordo con il contenuto di quell'antecedente. Inoltre, questa non è un'evenienza rara. Succede molte volte. Provate a chiederlo a chi abbia mai cercato di perdere peso, di smettere di fumare o di mantenere i suoi propositi per il nuovo anno.

Nella sua forma più cruda, l'errore che fa sembrare sconcertante il fatto che possa esserci l'*akrasia* deriva da una concezione errata della causazione. Se, per esempio, pensiamo alla causazione sulla base del modello delle palle da biliardo che colpiscono altre palle da biliardo o delle ruote dentate che mettono in moto altre ruote dentate, allora sembrerà semplicemente impossibile che noi possiamo avere delle cause senza effetti. Se le intenzioni causano il comportamento e l'intenzione era presente e l'agente non ha intrapreso l'azione intenzionale, questo può accadere solo perché altre cause hanno interferito, o perché non era il tipo d'intenzione che pensavamo fosse, o qualcosa di simile.

Ma la causazione intenzionale è sotto certi aspetti importanti differente dalla causazione del tipo "palle da biliardo". Entrambe sono casi di causazione, ma nel caso dei desideri e delle intenzioni, nel caso delle normali azioni volontarie, una volta che le cause sono presenti esse non costringono l'agente ad agire; l'agente deve *agire sulla base* delle ragioni o sulla base della sua intenzione. Nel caso dell'azione volontaria c'è, come abbiamo visto nei capitoli 1 e 3, un divario, una certa inerzia tra il processo di deliberazione e la formazione di un'intenzione, e c'è un altro divario tra l'intenzione e l'effettivo impegno.

Quando si parla di intenzionalità, è meglio pensare dal punto di vista della prima persona. Ebbene, che cosa significa per me formare un'intenzione e poi non agire in base ad essa? Mi viene sempre impedito di agire in base ad essa, e sono sempre costretto da delle cause, coscienti o non coscienti, ad agire contrariamente alla mia intenzione? Naturalmente no. Ebbene risulta sempre che in tali casi l'intenzione era in qualche modo in difetto, condizionata, o inappropriata in modo che essa non era un'intenzione fortissima,

incondizionata, priva di vincoli, ma solamente un'intenzione condizionata, *prima facie*? Ancora una volta naturalmente no. È possibile, come sappiamo, per un'intenzione essere forte e incondizionata quanto si vuole, senza nessuna interferenza, e che tuttavia l'azione non venga fatta.

Per vedere in che modo l'*akrasia* accada dobbiamo ricordarci di come si manifestano le azioni nei casi normali, nei casi privi di *akrasia*. Quando formo un'intenzione io devo ancora agire sulla base dell'intenzione che ho formato. Io non posso semplicemente sedermi e aspettare che l'azione accada, come faccio nel caso delle palle da biliardo. Ma dal punto di vista della prima persona, che è la sola prospettiva che qui ci interessa davvero, le azioni non sono semplicemente cose che accadono, esse non sono semplicemente eventi che avvengono; piuttosto, dal punto di vista della prima persona, esse vengono *fatte*; esse sono, per esempio, intraprese, iniziate o compiute. Decidersi non è sufficiente; si deve ancora "fare". È in questo divario tra intenzione e azione che noi troviamo la possibilità, o meglio l'inevitabilità di almeno alcuni casi di debolezza della volontà. A causa dell'inevitabilità dei desideri in conflitto e di altri motivatori, per gran parte delle azioni premeditate può esserci la possibilità che quando arrivi il momento di compiere l'azione l'agente si trovi di fronte al desiderio di non fare la cosa che aveva deciso di fare.

Che cosa succederebbe se l'*akrasia* fosse genuinamente impossibile? Immaginate un mondo in cui una volta che una persona ha formato un'intenzione incondizionata di compiere un'azione (ed ha soddisfatto tutte le altre condizioni antecedenti che si vogliano nominare, come il formare un giudizio di valore fortissimo in favore del compimento di essa, l'impartire un ordine morale a se stessi per compierla, ecc.), l'azione segua per necessità causale, a meno che qualche altra causa sovrasti il potere causale dell'intenzione o l'intenzione stessa si indebolisca e perda il suo potere di causare l'azione. Se questo fosse il modo in cui funzionasse il mondo noi non dovremmo *agire sulla base* delle nostre intenzioni; noi potremmo, per così dire, aspettare che esse agiscano da sole. Noi potremmo sederci e vedere come si sviluppano le cose. Ma noi non possiamo fare questo, noi dobbiamo sempre agire.

L'*akrasia*, è tutto tranne che un sintomo di un certo tipo di libertà, e capiremo meglio questo se esploreremo ulteriormente questa libertà. Secondo una certa concezione classica del prendere decisioni, noi da un momento all'altro, raggiungiamo un "punto di scelta": un punto in cui ci viene fornita una vasta gamma di opzioni dalla quale possiamo – o a volte

dobbiamo – scegliere. Contro questa concezione voglio osservare che in ogni momento normale, cosciente, di veglia nella nostra vita ci viene fornita una gamma di scelte indefinita, anzi, per la verità infinita. In questo momento, io sto scrivendo questo capitolo, posso muovere le dita dei piedi, muovere la mia mano sinistra, o la mia mano destra oppure partire per Timbuktu. L'esperienza di una qualsiasi azione libera, normale, cosciente contiene al suo interno la possibilità di non compiere quell'azione, ma di fare invece qualcos'altro. Molte di queste opzioni saranno impraticabili, in quanto sterili, non desiderabili o perfino ridicole. Ma tra le varie possibilità ve ne saranno sicuramente alcune che ci piacerebbe realmente fare, per esempio, bere un altro drink, andare a letto, andare a fare una passeggiata o semplicemente smettere di lavorare e leggere un romanzo.

Ci sono molte forme diverse di *akrasia*, ma il modo in cui l'*akrasia* normalmente insorge è questo: come risultato di una deliberazione noi formiamo un'intenzione. Ma dal momento che abbiamo sempre una gamma indefinita di scelte aperta dinanzi a noi, quando giunge il momento di agire sulla base dell'intenzione, molte delle altre scelte ci appaiono attraenti, o motivate su altre basi. Per la maggior parte delle azioni che compiamo per una ragione, vi sono ragioni per non compiere quell'azione ma fare invece qualcos'altro. Talvolta noi agiamo sulla base di quelle ragioni e non sulla base della nostra intenzione originaria. La soluzione al problema dell'*akrasia* è semplice: noi non abbiamo quasi mai una sola scelta a nostra disposizione. Indipendentemente da una particolare decisione, altre opzioni continuano ad essere attraenti.

Potrebbe allora sembrare sconcertante il fatto che noi agiamo sempre sulla base del nostro giudizio migliore, dal momento che su di noi premono tutte esigenze conflittuali. Ma non è così sconcertante se ripensiamo al perché abbiamo una deliberazione e delle intenzioni precedenti. Esse hanno essenzialmente la funzione di regolare il nostro comportamento. Il comportamento sano non è semplicemente un insieme di atti spontanei, ognuno motivato da considerazioni momentanee; piuttosto noi poniamo un ordine nella nostra vita e diventiamo capaci di soddisfare la maggior parte dei nostri obiettivi a lungo termine grazie alla formazione di intenzioni precedenti che avviene attraverso la deliberazione.

È cosa comune tracciare un'analogia tra l'*akrasia* e l'auto-inganno, e in effetti vi sono delle similitudini. Una forma caratteristica di *akrasia* è quella del dovere contrapposto al desiderio, proprio come una forma caratteristica di auto-inganno è l'evidenza

contrapposta al desiderio. Per esempio, l'innamorato inganna se stesso pensando che la sua amata gli sia fedele a dispetto della palese evidenza del contrario, poiché egli vuole disperatamente credere nella sua fedeltà. Ma vi sono alcune differenze cruciali tra l'*akrasia* e l'auto-inganno, e gran parte di queste hanno a che fare con la direzione di adattamento. La persona con una volontà debole può lasciare ogni cosa in superficie. Egli può dire a se stesso: "sì, lo so che non dovrei fumare un'altra sigaretta ed ho preso la ferma decisione di smettere, ma ugualmente desidero tantissimo fumarne una; e così, contro il mio miglior giudizio sto per fumarne una." Ma colui che si auto-inganna non può dire a se stesso: "sì, lo so che la proposizione in cui credo è certamente falsa, ma io desidero tantissimo crederla vera; e così, contro il mio miglior giudizio e conoscenza continuerò a credere ad essa." Una tale visione non è auto-inganno, è semplicemente irrazionale e forse persino incoerente. Per soddisfare il desiderio di credere ciò che sa essere falso, l'agente deve sopprimere la conoscenza. L'"*Akrasia*" indica un certo tipo di conflitto tra stati intenzionali, dove vince la parte sbagliata. L'"Auto-inganno" non indica un tipo di conflitto, ma è piuttosto una forma di annullamento dei conflitti per mezzo della soppressione della parte sgradita. È una forma di occultamento di un conflitto, in alcuni casi essa sarebbe in effetti un'incoerenza, che se venisse in superficie non potrebbe più essere mantenuto nascosto. La forma del conflitto è:

Io ho un'evidenza schiacciante che p (o addirittura, io so che p) ma io vorrei tantissimo credere che non p .

Se si manifesta in questa forma tale conflitto non può essere vinto dal desiderio. Perché il desiderio vinca, il conflitto stesso deve essere soppresso. Questo è il motivo per cui si tratta di un caso di auto-inganno. L'*akrasia* è una forma di conflitto ma non una forma di incoerenza logica o di contraddizione logica. L'auto-inganno è invece un modo di nascondere quella che sarebbe una forma di incoerenza o di contraddizione se ad essa fosse permesso di venire in superficie. Per queste ragioni l'auto-inganno richiede logicamente la nozione di inconscio; l'*akrasia* invece no. L'*akrasia* è spesso supportata dall'auto-inganno come un modo per rimuovere il conflitto, per esempio il fumatore dice a se stesso: "fumare non mi fa così male e, inoltre, la tesi secondo cui il fumo è la causa del cancro non è mai stata provata".

Per riassumere queste differenze: l'*akrasia* e l'auto-inganno non sono realmente simili nella struttura. L'*akrasia* ha tipicamente questa forma:

È meglio fare *A* e io ho deciso di fare *A*, ma sto volontariamente e intenzionalmente facendo *B*.

Non vi è qui alcuna assurdità logica o un'inconsistenza, sebbene vi sia un conflitto tra ragioni per l'azione inconsistenti, e sebbene l'atto sia irrazionale in quanto l'agente agisce intenzionalmente e volontariamente sulla base di una ragione che egli ritiene sia la ragione sbagliata per agire.

L'auto-inganno ha tipicamente questa forma:

L'agente ha lo stato cosciente: io credo che non *p*. Ed ha questi stati inconsci: ho una schiacciante evidenza che *p* e voglio tantissimo credere che non *p*.

L'auto-inganno implica l'irrazionalità dunque e in alcuni casi persino l'inconsistenza logica. Esso può esistere solamente se uno degli elementi è soppresso dalla coscienza.

Nelle azioni con volontà debole, il sé agisce sulla base di una ragione che quello stesso sé giudica non essere la migliore ragione per agire, e il sé agisce contro quella stessa ragione che ha giudicato come la migliore azione in base alla quale agire. Ci sono molte forme differenti che questo modello può assumere e molti gradi differenti di debolezza della volontà. Si è tentati di pensare che nei casi di debolezza della volontà il sé sia sopraffatto da qualche forte desiderio, in modo che il desiderio in base al quale il sé agisce fornisce una condizione causale sufficiente per agire. Senza dubbio vi sono casi del genere, ma essi non sono dei casi tipici. Nel caso tipico il divario esiste sia per l'azione con volontà debole che per l'azione con una volontà forte. Io ho bevuto un altro bicchiere di vino a dispetto del mio giudizio secondo il quale non avrei dovuto bere un altro bicchiere di vino. Ma il mio bere un bicchiere di vino non era frutto di un obbligo, di una costrizione o di una determinazione più di quanto lo fosse la mia azione basata sulla volontà forte quando ho agito in accordo con il mio migliore giudizio. Il divario è – o potrebbe essere – lo stesso in entrambi i tipi di casi. E questo è il motivo per cui l'atto della volontà debole è in questo senso irrazionale. È irrazionale da parte mia decidere genuinamente di fare la scelta sbagliata quando so che si

tratta della scelta sbagliata. La metafora della “debolezza” credo che in questi casi sia proprio giusta poiché il tema in questione riguarda il sé. Ciò che è in questione non riguarda la debolezza dei miei desideri o delle mie convinzioni ma riguarda la debolezza di me stesso nel portare a compimento le decisioni che ho preso.

In base alla spiegazione che ho presentato il problema della debolezza della volontà non è un problema serio in filosofia. Esso diviene serio solamente se presupponiamo un insieme errato di assunzioni circa gli antecedenti causali dell’azione. Ma è tuttavia illuminante poiché ci permette di osservare il divario da un diverso punto di vista. La domanda, comunque, rimane: qual è o quale potrebbe essere la realtà neurobiologica del divario? Questa è una domanda che rimando all’ultimo capitolo.

P53

8.

Perché non c'è una logica deduttiva della ragione pratica

I. La logica della ragione pratica

La ragione pratica, ci viene spesso detto, è il ragionare sul che cosa fare, e la ragione teorica è il ragionare sul che cosa credere. Ma se le cose stanno in questo modo, dovrebbe sembrarci sconcertante il fatto che non disponiamo di una spiegazione ampiamente accettata della struttura logica deduttiva della ragione pratica come invece sembriamo avere per la ragione teorica deduttiva. In effetti, i processi tramite i quali pensiamo al miglior modo possibile per raggiungere i nostri obiettivi sembrano essere dei processi razionali tanto quanto lo sono i processi tramite i quali pensiamo alle implicazioni delle nostre credenze; perché dovremmo dunque possedere una logica così potente per un tipo di ragione e non per l'altra? Aristotele inventò non soltanto il sillogismo teorico ma anche, sebbene questo sia stato meno influente, il sillogismo pratico. Perché non c'è dunque una teoria convenzionale del sillogismo pratico così come c'è una teoria convenzionale del sillogismo teorico e una teoria della logica deduttiva in generale?

Per comprendere qual è il problema, proviamo ad esaminare in che modo esso è stato apparentemente risolto per la ragione teorica. E' necessario innanzitutto distinguere le domande di relazione logica dalle domande della psicologia filosofica. Grandi progressi sono stati fatti nell'ambito della logica deduttiva quando, nel diciannovesimo secolo, Frege

ha separato le domande della psicologia filosofica (le “leggi del pensiero”) da quelle delle relazioni logiche. Dopo Frege, si è pensato che se si fossero interpretate in maniera corretta le relazioni logiche la psicologia filosofica avrebbe dovuto essere relativamente semplice. Per esempio, dopo aver compreso le relazioni di conseguenza logica tra le proposizioni molte delle corrispondenti domande riguardanti la credenza sembreranno piuttosto semplici. Se io so che le premesse “tutti gli uomini sono mortali” e “Socrate è un uomo”, insieme, portano alla conclusione “Socrate è mortale”, allora io so già che qualcuno che *crede* in quelle premesse è *vincolato* nei confronti di quella conclusione; che qualcuno che *sa* che quelle premesse sono vere è *giustificato* nell’*inferire* la verità della conclusione, ecc. Sembra, in breve, che vi sia un insieme piuttosto compatto di parallelismi, interno alla ragione teorica, tra nozioni “logiche” come la premessa, la conclusione e la consequenzialità logica da un lato e nozioni “psicologiche” come la credenza, l’impegno e l’inferenza dall’altro. La ragione di questo insieme compatto di parallelismi è che gli stati psicologici hanno dei contenuti proposizionali ed ereditano perciò alcune caratteristiche delle relazioni logiche tra le proposizioni. Poiché la consequenzialità logica conserva la verità, e la credenza è un impegno nei confronti della verità, le caratteristiche della consequenzialità logica possono essere tratteggiate sugli impegni della credenza. Se q è una conseguenza logica di p , e io credo che p , allora sono vincolato nei confronti della verità di q . Il principio implicito che ha funzionato così bene nella logica assertoria è che, se si hanno delle corrette relazioni logiche, allora gran parte della psicologia filosofica potrà funzionare autonomamente.

Ora, supponendo di accettare questa distinzione tra le relazioni logiche e la psicologia filosofica, in che modo si suppone che questo funzioni per la ragione pratica? Quali sono le relazioni logiche della ragione pratica e in che modo esse supportano la psicologia filosofica? Alcune delle domande riguardanti le relazioni logiche sarebbero: qual è la struttura logica formale dell’argomento pratico? In particolare, possiamo ottenere una definizione di validità formale per la ragione pratica così come nel caso della ragione “teorica” deduttiva? La logica pratica esibisce le stesse regole o richiede differenti regole di inferenza rispetto alla logica assertoria? Le domande circa la psicologia della deliberazione riguardano molti dei temi che abbiamo discusso in questo libro, specialmente il carattere degli stati intenzionali presenti nel ragionamento pratico, il loro relazionarsi alla struttura logica della deliberazione, le loro relazioni con l’azione e le loro relazioni con le ragioni per

l'azione in generale. Che tipi di stati intenzionali compaiono nella deliberazione e quali sono le relazioni tra di loro? Che genere di cose possono essere le ragioni per l'azione? Qual è la natura della motivazione e in che modo la deliberazione effettivamente motiva l'azione?

Alla luce della nostra distinzione tra teoria logica e psicologia filosofica, la domanda che stiamo ponendo è ora la seguente: “ci sono modelli formali della validità pratica, tali che l'*accettazione* delle premesse di un argomento pratico valido impegni qualcuno ad accettare la conclusione, nel modo che è caratteristico della ragione teorica?”. Abbiamo osservato che nella ragione teorica la credenza nelle premesse di un argomento valido impegna ad una credenza nelle conclusioni. Potremmo ottenere impegni simili nei confronti dei desideri e delle intenzioni in quanto conclusioni della ragione pratica? L'obiettivo di una logica formale della ragione pratica, mi sembra, dovrebbe essere quello di giungere ad un insieme di forme valide di inferenza pratica; e la verifica per un simile progetto è quella di vedere se l'agente che ha accettato le premesse di un'inferenza pratica valida sia impegnato a desiderare o intendere la conclusione, nello stesso modo in cui l'agente che accetta le premesse di un'inferenza teorica valida è impegnato a credere nelle conclusioni.

II. Tre modelli di ragione pratica

Vorrei iniziare, prendendo in considerazione alcuni tentativi di descrizione della struttura logica formale della ragione pratica. Limiterò la discussione al cosiddetto ragionamento mezzi-fini, poiché la maggior parte degli autori che hanno trattato l'argomento rientrano nella tradizione del modello classico e pensano che la ragione pratica sia una questione che riguarda la decisione dei mezzi necessari per raggiungere certi fini. Abbastanza stranamente, non è per niente facile o privo di controversie descrivere la struttura formale del ragionamento mezzi-fini, e non c'è un accordo generale su che cosa esso sia. Nella letteratura filosofica è presente una varietà sconcertante di modelli formali per tale ragionamento, e vi sono disaccordi fondamentali su quelli che si suppone siano i suoi elementi speciali – si tratta di desideri, intenzioni, comandi, imperativi, norme, *noemata*, azioni, o di che cosa?¹. Penso che la ragione di questa varietà sia che gli autori in

¹ Per una buona sintesi della letteratura relativa a questi temi, fino a metà degli anni '70, si veda Bruce Aune, *Reason and Action*, Dordrecht-Holland, D.Reidel Publishing Company, 1977, cap. 4, pp. 144-194.

questione si confrontano con il fatto che gli elementi del ragionamento sono dei fattivi e i fattivi possono presentarsi sotto forme differenti. Molti filosofi si sono espressi in maniera piuttosto vaga riguardo al modello di spiegazione credenze-desideri e alla deliberazione, ma quale si pensa sia esattamente la struttura di questo modello? Antony Kenny suggerisce che la struttura della ragione pratica è piuttosto differente da quella della ragione teorica. Egli fornisce il seguente esempio:

Devo essere a Londra per le 16.15

Se prendo quello delle 14.30 sarò a Londra per le 16.15

Quindi prenderò quello delle 14.30.²

Poiché le premesse esibiscono entrambe le direzioni di adattamento possiamo rappresentare la forma dell'argomento, in generale, con il seguente simbolismo, usando “↑” e “↓” rispettivamente per le direzioni di adattamento verso l'alto e verso il basso, e usando “F” e “M” per fini e mezzi:

↑ (F).

↓ (Se M allora F).

Perciò, ↑ (M).

Nel caso in cui si abbiano delle credenze e dei desideri come “premesse”, questo modello di inferenza può essere rappresentato come segue:

DES (io raggiungo F).

CRED (Se faccio M raggiungerò F).

Perciò, DES (io faccio M).

Ma questo non sembra essere corretto, poiché due premesse in questa forma semplicemente non impegnano una persona ad accettare la conclusione. Non vi è alcun impegno nei confronti di un desiderio, e ancor meno di un'intenzione, come conclusione di questa forma di argomento. Per vedere questo notate che molti dei fini F cui una persona

² Anthony Kenny, *Will, Freedom and Power*, New York, Barnes and Noble, 1976, p. 70.

può pensare sono piuttosto banali e molti mezzi *M* sono assurdi. Per esempio io vorrei che questa metropolitana fosse meno affollata e credo che se uccidessi tutti gli altri passeggeri sarebbe meno affollata. Ma questo non mi impegna a desiderare di uccidere gli altri passeggeri. Sicuramente si *potrebbe* formare un desiderio omicida in una metropolitana affollata, ma mi sembra assurdo sostenere che la razionalità mi *impegni* a desiderare di uccidere solo sulla base di altre mie credenze e desideri. Il massimo che questo modello potrebbe spiegare sarebbero le motivazioni *possibili* per formare un desiderio. Qualcuno che disponga delle credenze e dei desideri appropriati ha un motivo possibile per desiderare *M*. Ma non c'è alcun *impegno* nei confronti di un tale desiderio.

A volte si è detto che questo modello fallisce perché non c'è una relazione di implicazione tra i contenuti proposizionali delle premesse e della conclusione. Infatti, se guardiamo semplicemente ai contenuti proposizionali, l'inferenza è colpevole della fallacia di affermare il conseguente. Alcuni filosofi pensano che la forma tipica della ragione pratica vada cercata nei casi in cui il mezzo è una condizione necessaria per raggiungere il fine. Sostengono dunque quanto segue (o delle varianti di esso):

↑ (io raggiungo il fine *F*).

↓ (il solo modo per raggiungere *F* è per mezzo di *M*) (a volte formulato come “*M* è una condizione necessaria di *F*” oppure “per raggiungere *F*, devo fare *M*”).

Perciò, ↑ (io faccio *M*).

In questo caso la soddisfazione delle premesse garantisce la soddisfazione della conclusione, ma l'accettazione delle premesse non impegna ancora una persona a un desiderio o intenzione nella conclusione. Se si pensa a questo modello nei termini degli esempi della vita reale, esso appare improponibile come spiegazione generale della ragione pratica. In generale vi sono molti mezzi, molti dei quali assurdi, per raggiungere qualsiasi fine o scopo; e nei rari casi in cui ci sia un solo mezzo, questo potrebbe sembrare talmente assurdo da risultare completamente impossibile. Supponiamo che voi abbiate un fine a cui tenete: volete andare a Parigi, diventare ricchi o sposare un Repubblicano. Ebbene, nel caso di Parigi, per esempio, vi sono moltissimi modi per andarci. Potreste andare a piedi, a nuoto, in aereo, in nave, in kayak o con un razzo; potreste andarci facendo un tunnel attraverso la Terra, o passando dalla Luna o dal Polo Nord. In casi davvero molto rari può esserci un solo

un mezzo per un fine. Per quanto ne sappiamo non esiste alcun modo rapido per sbarazzarsi dei sintomi dell'influenza, se non la morte. Perciò, secondo il modello sopra esposto, se io desidero sbarazzarmi dei miei sintomi dell'influenza immediatamente, e credo che il solo modo per farlo sia la morte, io mi impegno a desiderare la mia morte. Questo modello, come il primo, ha pochissime applicazioni. Gran parte del ragionamento mezzi-fini non riguarda le condizioni necessarie, ed anche qualora le riguardasse, desiderare un certo fine non mi impegnerebbe a desiderare un certo mezzo.³

Nel primo di questi esempi non c'era una relazione di implicazione tra i contenuti proposizionali delle premesse e la conclusione; ma nel secondo tale relazione c'era. Il fatto che le relazioni di implicazione non generano un impegno nei confronti di un desiderio secondario rivela un importante contrasto tra la logica della sola credenza e la logica delle combinazioni credenze-desideri. Se credo sia che p che (se p allora q), allora sono impegnato a credere che q . Ma se io voglio p e credo che (se p allora q), non sono impegnato a volere q . Perché c'è questa differenza? Quando lo avremo capito, avremo fatto molti passi avanti nella comprensione del perché non esiste una logica plausibile della ragione pratica.

Proseguiamo ulteriormente nel nostro tentativo di costruire un modello logico formale della ragione pratica. Generalmente, quando si ha un desiderio, un'intenzione o un obiettivo non si cerca semplicemente un mezzo *qualsiasi*; né si cerca il *solo* mezzo possibile; si cerca il mezzo *migliore* (come dice Aristotele, si cerca il mezzo "migliore o più facile"). E se si è razionali, quando non sia disponibile alcun mezzo buono, o perlomeno ragionevole, si rinuncia completamente al fine che ci si era posti. Inoltre, poiché non si ha mai un solo obiettivo, se si è razionali, si valutano e selezionano i propri obiettivi alla luce di – ebbene, di che cosa? Dovremo ritornare su questo punto più avanti. Nel frattempo, supponiamo di aver selezionato seriamente un obiettivo e di ritenerlo un obiettivo ragionevole. Supponiamo di voler seriamente andare a Parigi, cioè, "di avere preso una decisione": si cerca dunque di immaginare il modo migliore per arrivarci e si conclude che il modo migliore è costituito dall'aereo. C'è un modello formale plausibile della logica del ragionamento mezzi-fini per un caso del genere?

La forma dell'argomentazione sembra essere questa:

³ Aune, *Reason and Action*, il quale osserva che il primo modello è inadeguato per ragioni simili a quelle che ho suggerito io, ciononostante non riesce a vedere che lo stesso tipo di obiezioni si possono applicare al secondo modello.

Des (io vado a Parigi).

Cred (il miglior modo, tutto considerato, è andare in aereo).

Perciò, Des (io vado in aereo).

Se separiamo le questioni delle relazioni logiche dalle questioni della psicologia filosofica – come ho cercato di sostenere – notiamo che da un punto di vista logico questa argomentazione, in questa forma, è entimematica. Per essere formalmente valida richiederebbe un'ulteriore premessa del tipo:

Des (se vado a Parigi, vado nel modo migliore, tutto considerato).

Se aggiungiamo questa premessa, l'argomento è valido secondo i criteri della logica classica. Siano P = io vado a Parigi, Q = ci vado nel modo migliore e R = io vado in aereo. La sua forma, allora, sarà:

P

$P \rightarrow Q$

$Q \Leftrightarrow R$

$\therefore R$

E sebbene l'argomento non preservi la verità perché due delle sue premesse e la sua conclusione non hanno un valore di verità, questo non ha grande importanza poiché l'argomento preserva la soddisfazione, e la verità è semplicemente un caso speciale di soddisfazione. La verità è la soddisfazione di rappresentazioni con direzione di adattamento parola-a-mondo.

Ma, ancora una volta, come negli esempi precedenti, sembra che le relazioni logiche non esprimano la psicologia filosofica nel modo corretto. È senza dubbio evidente che una persona razionale che possieda tutte queste premesse debba avere, o debba essere impegnata ad avere, il desiderio di andare in aereo. Inoltre, per rendere questo plausibile, dobbiamo introdurre una premessa equivoca riguardante il voler fare le cose “nel modo migliore, tutto

considerato”. Sembra in effetti che qualsiasi tentativo di esprimere formalmente la struttura di un argomento pratico di questo tipo richieda, in generale, questa premessa, ma non è del tutto chiaro che cosa essa significhi. Che cosa si intende con “modo migliore” e con “tutto considerato”? Si noti, inoltre, che tali premesse non hanno un analogo nei casi tipici della ragione teorica. Quando si ragiona partendo dalle credenze secondo cui tutti gli uomini sono mortali, e che Socrate è un uomo, fino alla conclusione che Socrate è mortale, non si ha bisogno di nessuna premessa che riguardi qual è la cosa migliore a cui credere, tutto considerato.

Fino ad ora ho cercato di costruire un modello logico formale della concezione tradizionale del ragionamento mezzi-fini (concezione che si rifà ad Aristotele), e mi sembra che questo sia quanto di meglio potevo raggiungere. Ho anche tentato di fornire un’esposizione della sua struttura formale che potrebbe costituire un miglioramento delle versioni precedentemente proposte. Ma credo che tutto questo sia ancora disperatamente inadeguato. Dopo vari tentativi senza successo sono arrivato, se pure con riluttanza, alla conclusione che sia impossibile ottenere una logica formale della ragione pratica che sia adeguata ai fatti della psicologia filosofica. Per mostrare perché le cose stanno in questo modo, mi rivolgerò ora alla discussione della natura del desiderio. La caratteristica essenziale per la presente discussione è che essa ha una direzione di adattamento verso l’alto. Molte delle caratteristiche che specificherò come caratteristiche del desiderio sono anche caratteristiche di altri fattivi con direzione di adattamento verso l’alto come gli obblighi, i bisogni, gli impegni ecc. Tuttavia, per semplicità, tratterò gran parte della discussione in termini di desiderio e la amplierò agli altri fattivi con direzione di adattamento verso l’alto solo più avanti.

III. La struttura del desiderio

Per comprendere le debolezze presenti nella mia revisione logica del ragionamento pratico, e per comprendere gli ostacoli generali nei confronti di una logica formale del ragionamento pratico, dobbiamo esplorare alcune caratteristiche generali del desiderio e, specialmente, esplorare le differenze tra i desideri e le credenze. Userò la spiegazione generale dell’intenzionalità che ho fornito nel capitolo 2, e le altre caratteristiche della teoria

dell'intenzionalità che ho presentato nel libro omonimo⁴. Specificamente, assumerò che, contrariamente alla grammatica di superficie delle espressioni di desiderio, tutti i desideri hanno intente proposizioni come contenuti intenzionali (così, “io voglio la tua automobile” significa qualcosa del tipo “io voglio avere la tua automobile per me”); quei desideri hanno una direzione di adattamento mondo-a-mente, mentre le credenze hanno direzione di adattamento mente-a-mondo; e quei desideri non hanno alcuna restrizione del contenuto intenzionale come hanno invece le intenzioni. Le intenzioni devono riguardare le azioni future o presenti dell'agente e devono avere un'autoreferenzialità causale insita nel loro contenuto intenzionale. I desideri non hanno tali condizioni causali, e possono riguardare qualsiasi cosa, il passato, il presente o il futuro. Sosterrò inoltre che le spiegazioni correnti della distinzione *de re/de dicto* sono desolatamente confuse, come lo è la tesi per cui i desideri sono intensionali-con-la-s. La distinzione *de re/de dicto* è propriamente costruita come una distinzione tra tipi di *enunciati* riguardanti i desideri, non tra tipi di desideri. La tesi secondo la quale tutti i desideri, le credenze ecc. sono, in genere, intensionali è semplicemente falsa. *Gli enunciati riguardanti* i desideri, le credenze ecc. sono, in genere, intensionali. Ma i desideri e le credenze non sono, in generale, essi stessi intensionali, sebbene in taluni casi bizzarri possano esserlo.⁵

Quando si desidera un certo stato di cose per soddisfare qualche altro desiderio, è meglio ricordare che ogni desiderio è parte di un desiderio più ampio. Se voglio andare nel mio ufficio per prendere la posta, c'è in effetti un desiderio il cui contenuto è semplicemente: io voglio che (io vado nel mio ufficio). Ma questo è parte di un desiderio più ampio il cui contenuto è: io voglio che (io prendo la mia posta andando nel mio ufficio). Questa caratteristica è condivisa dalle intenzioni. Se io intendo fare *a* per fare *b*, allora ho un'intenzione complessa la cui forma è: io intendo (io faccio *b* per mezzo del fare *a*). Mi soffermerò ulteriormente su questo punto più in avanti.

La prima caratteristica del desiderare (volere, augurarsi, ecc.), che lo differenzia dalla credenza, è che è possibile per un agente, in modo coerente e consapevole, volere che *p* e volere che non *p* così come per lui non è possibile, in modo coerente e consapevole, credere che *p* e credere che non *p*. E questa tesi è più forte della tesi secondo cui un agente può

⁴ John R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press, 1983, tr. it. *Della intenzionalità: un saggio di filosofia della mente*, Bompiani, Milano 1985.

⁵ Per la discussione di questi punti riguardanti l'intensionalità-con-la-s e la distinzione *de re/de dicto*, si veda Searle, *Intentionality*, capp. 7 e 8.

coerentemente avere dei desideri che è impossibile soddisfare simultaneamente perché egli non conosce le caratteristiche di tali desideri. Per esempio, Edipo desidera sposare una donna che ricada sotto la descrizione di “la mia fidanzata” e non desidera sposare nessuna donna che ricada sotto la descrizione di “mia madre” anche se di fatto vi è una donna che soddisfa entrambe le descrizioni. Ma io sto sostenendo che egli può, coerentemente, voler sposare Giocasta e non voler sposare Giocasta, sotto la medesima descrizione. I casi usuali di questo genere sono i casi in cui egli ha certe ragioni per volerla sposare e altre ragioni per non volerla sposare. Per esempio egli potrebbe volerla sposare – perché, diciamo, la trova bella e intelligente e, allo stesso tempo, non la vuole sposare – perché, diciamo, lei russa e si schiocca le dita. Questi casi sono comuni, ma è anche importante sottolineare che una persona potrebbe trovare le stesse caratteristiche allo stesso tempo desiderabili e indesiderabili. Egli potrebbe trovare la sua bellezza e la sua intelligenza esasperanti, e insieme attraenti, ed egli potrebbe trovare il suo russare e il suo schioccarsi le dita delle abitudini tanto tenere quanto insopportabili. (Immaginate che egli pensi tra sé: “è meraviglioso che lei sia così bella e intelligente, ma allo stesso tempo è un po’ stancante; quel suo stare seduta lì ed essere bella e intelligente tutto il giorno. Ed è esasperante sentirla russare e schioccarsi le dita, ma allo stesso tempo c’è qualcosa di tenero in questo. È così umano.”). Questa è la condizione umana.

La possibilità di sostenere razionalmente e coerentemente dei desideri incoerenti ha la spiacevole conseguenza logica che il desiderio non è racchiuso nella congiunzione. Così, se io desidero che p e desidero che non p , da questo non segue che io desidero che (p e non p). Per esempio io voglio essere a Berkeley proprio ora e voglio essere a Parigi proprio ora, ma sapendo che questi sono dei desideri inconsistenti, non si dà il caso che io mi impegni razionalmente a volere (essere simultaneamente a Berkeley e a Parigi proprio ora).

Per comprendere la possibilità di sostenere dei desideri inconsistenti, in modo razionale e coerente, e le sue conseguenze per la ragione pratica, è necessaria un’indagine più profonda. È consueto, e credo generalmente corretto distinguere, così come fa il modello classico, tra desideri primari e desideri secondari o derivati. È letteralmente vero dire al mio agente di viaggio “vorrei acquistare un biglietto aereo”. Ma io non ho una bramosia, un desiderio, una gran voglia o una passione per i biglietti aerei – essi sono semplicemente dei “mezzi” per dei “fini”. Un desiderio che è primario in relazione ad un certo desiderio può essere secondario in relazione ad un altro. Il mio desiderio di andare a Parigi è primario

relativamente al mio desiderio di acquistare un biglietto aereo, e secondario relativamente al mio desiderio di visitare il Louvre. La distinzione dei desideri in primario e secondario sarà allora sempre relativa a qualche struttura tramite la quale un desiderio è motivato da un altro desiderio o da qualche altro motivatore. Questa è esattamente l'immagine che è incorporata nella concezione classica della ragione pratica. In tali casi, come ho sottolineato, la completa specificazione del desiderio secondario fa riferimento al desiderio primario. Io non voglio semplicemente acquistare un biglietto, io voglio acquistare un biglietto per andare a Parigi.

Una volta che abbiamo compreso la caratteristica dei desideri secondari possiamo vedere che ci sono almeno due modi in cui gli agenti razionali possono formare dei desideri in conflitto tra loro. Per prima cosa, come abbiamo notato prima, un agente può semplicemente avere delle inclinazioni in conflitto tra loro. Ma, in secondo luogo, egli può formare dei desideri in conflitto tra loro da un insieme coerente di desideri primari unito alle credenze riguardanti il mezzo migliore per soddisfarli. Consideriamo l'esempio dell'uomo che ragiona sul fatto di voler andare a Parigi in aereo. Quest'uomo ha un desiderio secondario di andare in aereo, motivato da un desiderio di andare a Parigi, insieme con una credenza secondo la quale il modo migliore per andarci è l'aereo. Ma lo stesso uomo potrebbe avere costruito la seguente inferenza pratica: io non voglio fare niente che mi provochi la nausea o che mi terrorizzi, ma andare da qualsiasi parte in aereo mi provoca la nausea e mi terrorizza, perciò io non voglio andare da nessuna parte in aereo, quindi io non voglio andare a Parigi in aereo. È abbastanza facile sostenere questo secondo il modello del ragionamento pratico che ho suggerito sopra: tutto considerato il modo migliore per soddisfare il mio desiderio di evitare la nausea e il terrore è di *non* andare a Parigi in aereo. Dal momento che questo può essere affermato in quanto parte del ragionamento pratico, sembra che *la stessa persona, usando due ragionamenti indipendenti della ragione pratica, possa formare razionalmente dei desideri secondari incoerenti pur partendo da un'insieme coerente di credenze effettive e da un insieme coerente di desideri primari*. Un insieme coerente di “premesse” genererà dei desideri secondari incoerenti come “conclusioni”. Questa non è una caratteristica paradossale o incidentale del ragionare sulla base di credenze e desideri; piuttosto, è la conseguenza di certe differenze essenziali tra la ragione pratica e la ragione teorica.

Esaminiamo ulteriormente queste differenze: in generale è impossibile avere un insieme qualsiasi di desideri, persino un insieme coerente di desideri primari, senza avere, o

almeno senza essere razionalmente motivati ad avere, dei desideri incoerenti. O, per specificare meglio questo punto: se si prende l'insieme dei desideri e delle credenze di cui dispone una persona in un qualsiasi momento della sua vita, e si calcola quali desideri secondari possono essere razionalmente motivati dai suoi desideri primari, assumendo la verità delle sue credenze, si troveranno dei desideri incoerenti. Io non so in che modo dimostrarlo, ma un ampio numero di esempi potrebbe essere usato per illustrare questo fatto. Consideriamo l'esempio di andare a Parigi in aereo. Anche se gli aerei non mi provocano nausea e non mi terrorizzano, tuttavia io non voglio spendere del denaro; io non voglio rimanere seduto in aereo; non voglio mangiare il cibo che viene dato in aereo; non voglio fare la coda in aeroporto; non voglio sedermi vicino a persone che mettono il loro ginocchio proprio dove sto cercando di mettere il mio. E in effetti, io non voglio fare tutta una quantità di altre cose che sono il prezzo, in senso sia letterale che figurato, del soddisfare il mio desiderio di andare a Parigi in aereo. Lo stesso tipo di ragionamento che può portarmi a formare il desiderio di andare a Parigi in aereo potrebbe anche portarmi a formare il desiderio di *non* andare a Parigi in aereo.

Una risposta possibile a questo, implicita in almeno una parte della letteratura, è data dal richiamo alla nozione di preferenza. Io preferisco andare a Parigi in aereo e stare scomodo piuttosto che non andare a Parigi in aereo e stare comodo. Ma questa risposta, sebbene sia in parte accettabile, implica erroneamente che le preferenze siano date *precedentemente* al ragionamento pratico, poiché esse sono il prodotto caratterizzante del ragionamento pratico. E dal momento che le preferenze ordinate sono caratteristicamente prodotte dalla ragione pratica, esse non possono essere trattate come il suo presupposto universale. Come è un errore supporre che una persona razionale debba avere un insieme consistente di desideri, così è un errore supporre che le persone razionali debbano avere un insieme ordinato di (combinazioni di) desideri che precedono la deliberazione.

Questo porta alla seguente conclusione: anche se noi limitassimo la nostra discussione sul ragionamento pratico ai casi del tipo mezzi-fini, risulterebbe che la ragione pratica implica essenzialmente il giudizio di desideri in conflitto e di altri tipi di motivazioni in conflitto (per esempio, i fattivi con direzione di adattamento verso l'alto) mentre la ragione teorica non implica essenzialmente il giudizio di credenze in conflitto. La ragione pratica riguarda caratteristicamente il giudicare tra desideri in conflitto, obblighi, impegni, bisogni, richieste e doveri. Questo è il motivo per cui nel nostro tentativo di fornire una

spiegazione plausibile della concezione classica dell'inferenza pratica avevamo bisogno di un passaggio che riguardasse il volere andare con "il mezzo migliore, tutto considerato". Questo passaggio è caratteristico di qualsiasi ricostruzione razionale di un processo di ragione mezzi-fini, poiché "migliore" significa semplicemente quello che meglio riconcilia tutti i desideri in conflitto e altri motivatori che hanno attinenza con il caso. Comunque, questo ha anche la conseguenza che la formalizzazione della concezione classica che ho fornito è essenzialmente una banalizzazione del problema, poiché la parte difficile non è stata analizzata: come arriviamo alla conclusione che proprio quello è "il miglior modo di fare qualcosa tutto considerato"? E come riconciliamo le conclusioni inconsistenti degli insiemi in competizione di tali derivazioni valide?

Se dovessimo proseguire basandoci soltanto sulla concezione classica del ragionamento per mezzi e fini, allora per raggiungere la conclusione dell'argomento che potrebbe formare la base dell'azione, si dovrebbe passare per un intero insieme di altre simili catene inferenziali e poi trovare un modo per risolvere il problema delle ragioni in conflitto. *La concezione classica funziona in base al principio corretto secondo il quale qualsiasi mezzo per un fine desiderabile è desiderabile almeno fino a che conduce a quel fine. Ma il problema è che nella vita reale qualsiasi mezzo può essere, e generalmente sarà, non desiderabile per una serie di altri motivi, e il modello non ha modo di dimostrare in che modo questi conflitti vengono giudicati.*

La questione sembra immediatamente peggiorare quando consideriamo un'altra caratteristica dei desideri che avevamo già notato. Una persona che crede che p e che (se p allora q) è impegnata nei confronti della verità di q ; ma una persona che desidera che p e crede che (se p allora q) non è impegnata a desiderare che q .⁶ Si può volere che p e credere che (se p allora q) senza essere impegnati a volere che q . Per esempio non c'è nulla di logicamente sbagliato in una coppia che voglia avere un rapporto sessuale e che crede che se lo avessero lei rimarrebbe incinta ma non vogliono che lei rimanga incinta.

Possiamo riassumere questi punti riguardanti il desiderio e la distinzione tra il desiderio e la credenza nel modo seguente: i desideri hanno due caratteristiche speciali che

⁶ Certamente non si è impegnati nei confronti di una credenza nel senso che si deve effettivamente aver formato la credenza che q . Si potrebbe credere che p e che (se p allora q) senza averci più pensato. (Qualcuno potrebbe credere che 29 è un numero dispari e che non è divisibile per 3, 5, 7 o 9 e che qualsiasi numero che soddisfa queste condizioni è un numero primo, senza aver mai effettivamente tracciato la conclusione, per esempio aver formato la credenza che 29 è un numero primo).

rendono impossibile avere una logica formale della ragione pratica che sia parallela a quella che supponiamo essere la logica formale della ragione teorica. Potremmo definire la prima caratteristica “la necessità dell’inconsistenza”. Qualsiasi essere razionale nella vita reale è costretto ad avere desideri inconsistenti e altri tipi di motivatori. Potremmo definire la seconda caratteristica “la non distaccabilità del desiderio”. Insieme di credenze e desideri come “premesse” non vincolano necessariamente l’agente ad avere un desiderio corrispondente come “conclusione” persino in casi in cui i contenuti proposizionali delle premesse implicano il contenuto proposizionale della conclusione. Queste due tesi ci spiegano perché nella letteratura filosofica non vi sia alcuna spiegazione lontanamente plausibile di una struttura logica deduttiva della ragione pratica.

La morale è questa: per quel che posso dire, la ricerca di una struttura logica deduttiva formale della ragione pratica è fuorviante. I modelli che sono stati elaborati hanno pochissime, o addirittura nessuna, applicazione, e, se vengono predisposti per essere applicati nella vita reale, questo accade soltanto tramite una banalizzazione della caratteristica essenziale della deliberazione pratica: la riconciliazione tra desideri conflittuali e ragioni per l’azione conflittuali e la formazione di desideri razionali sulla base di tale riconciliazione. Potremmo sempre costruire un modello deduttivo per qualsiasi parte del ragionamento; ma quando una caratteristica essenziale del ragionamento contiene sia p che non p – come quando io voglio che p e voglio che non p , o sono obbligato a provocare che p e obbligato a provocare che non p – la logica deduttiva non è illuminante perché non può far fronte a tali incoerenze. I modelli devono sostenere o che le incoerenze non esistono o che esse sono state risolte (“nel modo migliore tutto considerato”). La prima via è quella intrapresa dai modelli che ho criticato all’inizio, la seconda via è quella intrapresa dalla mia versione modificata. La possibilità, per la verità l’inevitabilità, di desideri, obblighi, bisogni, ecc. contraddittori rende la concezione classica poco illuminante come modello per la struttura della deliberazione. Inoltre anche se si imbrogliasse fino al punto di banalizzare il problema non si avrebbe ancora un impegno nei confronti un desiderio come conclusione dell’argomento. Il *modus ponens* semplicemente non funziona nel caso della combinazione desiderio/credenza nel produrre un impegno a desiderare la conclusione.

Potrebbe il *modus ponens* funzionare per le combinazioni di desiderio/credenza? Questo non è un argomento tipico del ragionamento per mezzi e fini, ma vale la pena considerare la questione. Mi sembra che se si vuole che p e che se p allora q , si è impegnati a volere che q ,

ma si potrebbe anche razionalmente volere che non q . Così io potrei voler essere molto ricco e, coerentemente alla politica pubblica, vorrei essere molto ricco per essere tassato pesantemente, e logicamente parlando questo mi vincola a desiderare che se io divento ricco devo essere tassato pesantemente. Io sono dunque impegnato nei confronti di un tale desiderio, ma allo stesso tempo non vorrei essere tassato pesantemente. Così io ho un impegno verso un desiderio che è incoerente con un altro mio desiderio.

IV. La spiegazione delle differenze tra desiderio e credenza

Ora perché dovrebbero esserci queste differenze? Che cosa c'è nella psicologia del desiderio che lo rende logicamente così diverso dalla credenza? Ebbene, ogni risposta a questo risulta tautologica e quindi insoddisfacente, ma tentiamo ugualmente di esaminarla.

Sia i desideri che le credenze hanno dei contenuti proposizionali, entrambi hanno una direzione di adattamento, entrambi rappresentano le loro condizioni di soddisfazione, ed entrambi rappresentano le loro condizioni di soddisfazione sotto certi aspetti. Qual è dunque la differenza che spiega le diverse proprietà logiche dei desideri e delle credenze? La differenza deriva da due caratteristiche correlate, la differenza nella direzione di adattamento e la differenza nell'impegno. Il compito delle credenze è rappresentare come stanno le cose (direzione di adattamento verso il basso) e colui che ha una credenza è impegnato nei confronti della sua verità. Nella misura in cui la credenza riesce in questo o non vi riesce, essa sarà, rispettivamente, vera o falsa. Il compito dei desideri non è di rappresentare come le cose sono, ma come vorremmo che esse fossero. E i desideri possono riuscire a rappresentare come vorremmo che le cose fossero anche quando le cose non risultano essere nel modo in cui noi volevamo che esse fossero. Nel caso della credenza, il contenuto proposizionale rappresenta un certo stato di cose come effettivamente esistenti. Ma nel caso del desiderio, il contenuto proposizionale non funziona in modo da rappresentare un effettivo stato di cose, ma piuttosto uno stato di cose *desiderato*, che potrebbe essere reale, non esistente, possibile, impossibile o ciò che si vuole. E il contenuto proposizionale rappresenta lo stato di cose sotto gli aspetti che l'agente trova desiderabili. Non c'è niente di sbagliato nei desideri insoddisfatti, in quanto desideri, mentre c'è qualcosa di sbagliato nelle credenze insoddisfatte, in quanto credenze, e cioè che sono false. Esse

falliscono nel loro compito di rappresentare il modo in cui le cose sono. I desideri riescono nel loro compito di rappresentare il modo in cui vorremmo che le cose fossero anche nei casi in cui le cose non sono nel modo in cui noi vorremmo che esse fossero, e cioè perfino in casi in cui le loro condizioni di successo non sono soddisfatte. Parlando in modo grossolano, quando la mia credenza è falsa, è la credenza che è in errore. Quando il mio desiderio è insoddisfatto è il mondo ad essere in errore.

Le due caratteristiche logiche del desiderio, l'incoerenza e la non distaccabilità, derivano entrambe da questa caratteristica sottostante: i desideri sono inclinazioni verso stati di cose (possibili, reali o impossibili) sotto certi aspetti. Non c'è un'irrazionalità necessaria implicata dal fatto che si può essere inclini o non inclini nei confronti dello stesso stato di cose sotto lo stesso aspetto; e il fatto che si sia inclini verso uno stato di cose sotto un certo aspetto assieme alla conoscenza delle conseguenze dell'esistenza di quello stato di cose non garantisce che, se razionali, si sarebbe inclini a quelle conseguenze.

Ma se si prova a sostenere la medesima cosa riguardo alla credenza, questo non funziona. Le credenze sono convinzioni che gli stati di cose esistono sotto certi aspetti. Ma non si può razionalmente essere convinti che uno stato di cose esista e non esista sotto il medesimo aspetto. E il fatto che si sia convinti dell'esistenza di uno stato di cose sotto un certo aspetto assieme alla conoscenza delle conseguenze dell'esistenza di quello stato di cose garantisce che, se si è razionali, si sarà convinti (o almeno vincolati nei confronti) di quelle conseguenze. È importante sottolineare che queste caratteristiche della credenza provengono da due delle sue caratteristiche: la direzione di adattamento verso il basso e il vincolo, o impegno. Se si ha solamente la direzione di adattamento verso il basso, questo non è sufficiente. Così le ipotesi possibili riguardanti come le cose potrebbero essere devono avere anche la direzione di adattamento verso il basso. Ma se è possibile concepire in modo coerente e razionale delle ipotesi inconsistenti non è invece possibile sostenere in modo coerente e razionale delle credenze inconsistenti, e questo perché le credenze, a differenza delle ipotesi, sebbene entrambe implicino la direzione di adattamento verso il basso, hanno la caratteristica aggiuntiva del vincolo.

Queste caratteristiche del desiderio sono caratteristiche di altri tipi di rappresentazioni con direzione di adattamento mondo-a-parola. Le caratteristiche della incoerenza, o inconsistenza, e della non distaccabilità si applicano ai bisogni e agli obblighi così come ai desideri. Io posso avere coerentemente dei bisogni e degli obblighi

inconsistenti e non avere necessariamente bisogno delle conseguenze dei miei bisogni, né sono obbligato a realizzare le conseguenze dei miei obblighi. Esempi di questi fenomeni non sono difficili da trovare: io potrei aver bisogno di prendere una medicina per alleviare un certo insieme di sintomi, ma ho bisogno di evitare quella medicina, perché aggrava un altro insieme di sintomi. Io ho l'obbligo di incontrare la mia classe all'università, ma ho anche l'obbligo di fare lezione in un'altra università, perché l'ho promesso un anno fa. Ho bisogno di prendere l'aspirina per evitare un disturbo cardiaco, ma l'aspirina danneggia il mio stomaco, e quindi ho bisogno di evitare l'aspirina. Jones ha l'obbligo di sposare Smith perché lei ha fatto una promessa, ma sposando Smith renderà i suoi genitori infelici, e lei non ha l'obbligo di rendere i suoi genitori infelici. È sorprendente, a proposito, quanto l'opacità referenziale di tutti questi concetti, "obbligo", "bisogno", ecc. sia trascurata nella letteratura.

Nell'obbiettare a questa spiegazione, si potrebbe dire, "guarda, quando credo qualcosa, ciò che credo è che sia vera. Così se io credo qualcosa e so che essa non può essere vera a meno che qualcos'altro non sia vero, allora la mia credenza e conoscenza devono impegnarmi anche nei confronti della verità di quell'altra cosa. Ma perché non accade la stessa cosa con il desiderio? Quando voglio qualcosa ciò che voglio è che quel qualcosa accada o esista, ma se so che non può accadere o che non può esistere a meno che non accada qualcos'altro o esista qualcos'altro allora sicuramente io devo essere impegnato a volere quel qualcos'altro". Ma l'analogia crolla. Se io voglio trapanarti il dente per riempire la cavità e so che trapanando il dente causerò del dolore, da questo non segue il fatto che io sia in qualche modo impegnato a causare del dolore, e ancor meno a voler causare del dolore. E la dimostrazione di questa distinzione è piuttosto semplice: se io fallisco nel causare dolore una delle mie credenze è perciò falsa, ma nessuno dei miei desideri è, per questo stesso motivo, non soddisfatto.

Quando voglio qualcosa, la voglio soltanto sotto certi aspetti. "Sì, ma anche quando io credo qualcosa la credo solo sotto certi aspetti. Gli enunciati di credenza sono semplicemente opachi tanto quanto lo sono gli enunciati di desiderio". Sì, ma c'è questa differenza: quando qualcosa è desiderato sotto certi aspetti sono, in generale, quegli aspetti che lo rendono desiderabile. Infatti le relazioni tra gli aspetti e le ragioni per desiderare sono piuttosto differenti dal caso della credenza, dal momento che *la specificazione delle ragioni per desiderare qualcosa è, in generale, già una specificazione del contenuto del desiderio;*

ma la specificazione dell'evidenza sulla base della quale io sostengo una credenza non è, in generale, essa stessa parte della specificazione della credenza. La relazione che c'è tra le ragioni per credere e le proposizioni credute è differente dalla relazione che c'è tra i contenuti delle ragioni per voler fare e la proposizione che è il contenuto del desiderio, poiché, in generale, gli enunciati delle ragioni per volere esprimono una parte di ciò che si vuole. Se si vuole qualcosa per una ragione allora la ragione è parte del contenuto del nostro desiderio. Per esempio, se voglio che piovva perché il mio giardino cresca, allora io desidero sia che piovva sia che il mio giardino cresca. Se io credo che pioverà e credo che la pioggia farà crescere il mio giardino io credo sia che pioverà sia che il mio giardino crescerà. Ma c'è ancora una differenza cruciale. Se io voglio che piovva *per far sì che* il mio giardino cresca, allora la mia ragione per volere che piovva è parte dell'intero contenuto dell'intero desiderio complesso. La mia ragione per credere sia che pioverà sia che la pioggia farà crescere il mio giardino, d'altro canto, ha a che fare con un vasto numero di evidenze che riguardano la meteorologia, l'affidabilità delle previsioni del tempo, e la funzione dell'umidità nel produrre la crescita delle piante. Tutte queste considerazioni valgono come evidenza per la verità della mia credenza, ma non sono esse stesse il contenuto proprio di quella credenza. Ma nel caso del mio desiderio, il ruolo delle ragioni non è per nulla quello dell'evidenza, poiché le ragioni indicano gli aspetti sotto i quali il fenomeno in questione è desiderato. Le ragioni, in breve, sono parte del contenuto del desiderio complesso.

Per riassumere: le credenze hanno una direzione di adattamento mente-a-mondo, e colui che ha la credenza è impegnato nei confronti di un adattamento effettivamente esistente e cioè, è impegnato nei confronti della verità della credenza. I desideri hanno una direzione di adattamento mondo-a-mente e colui che ha un desiderio non ha bisogno di essere impegnato nei confronti della sua soddisfazione. Il compito del desiderio non è di rappresentare come stanno le cose, ma come vorremmo che esse stessero. È la nozione di "impegno nei confronti di come stanno le cose" che blocca la semplice possibilità di sostenere coscientemente delle credenze contraddittorie, e che richiede un impegno nei confronti delle conseguenze delle proprie credenze, ma non esiste un tale blocco né una tale richiesta quando si tratta di come vorremmo che le cose fossero. Nonostante certe somiglianze formali, la credenza è realmente radicalmente diversa dal desiderio sia dal punto di vista delle sue caratteristiche logiche che di quelle fenomenologiche.

Per queste ragioni, è fuorviante pensare alla ragione teorica come ad un ragionamento che riguarda il che cosa credere nello stesso modo in cui pensiamo alla ragione pratica come ad un ragionamento che riguarda il che cosa fare. Ciò che si dovrebbe credere dipende dai singoli casi esistenti. Il ragionamento teorico riguarda dunque ciò che si deve credere solo in senso derivativo. Esso riguarda primariamente ciò che esiste – ciò che deve esistere date certe premesse. Inoltre, ora possiamo vedere che è fuorviante pensare che ci sia una “logica” della ragione teorica. C’è semplicemente la logica – che ha che fare, per esempio, con le relazioni logiche tra le proposizioni. La logica ci dice molto di più riguardo alla struttura razionale della ragione teorica che non riguardo alla struttura razionale della ragione pratica, poiché c’è una stretta connessione tra i vincoli razionali sulla credenza e le relazioni logiche tra le proposizioni. Questa connessione deriva dal fatto che, ripeto, le credenze sono destinate ad essere vere. Ma non c’è una tale stretta connessione tra la struttura del desiderio e la struttura della logica. A causa della direzione di adattamento verso l’alto dei desideri, io posso, e devo avere desideri in conflitto tra loro anche quando sono presenti tutti i fatti.

V. Alcune caratteristiche speciali delle intenzioni

Mi sono concentrato sui desideri, ma le intenzioni sono, sotto aspetti importanti, differenti dai desideri. Come i desideri, le intenzioni hanno una direzione di adattamento verso l’alto ma, diversamente dai desideri, riguardano sempre l’agente come soggetto e sono causalmente autoreferenziali. La mia intenzione è realizzata solo se agisco realizzando le mie intenzioni. Per questa ragione le intenzioni hanno un vincolo logico piuttosto diverso dal desiderio. È logicamente inconsistente avere delle intenzioni incoerenti mentre non è logicamente inconsistente avere dei desideri incoerenti. Le intenzioni sono progettate per causare le azioni e per questa ragione non possono funzionare se sono inconsistenti. Questa proibizione nei confronti dell’inconsistenza è condivisa da altri motivatori autoreferenziali come gli ordini e le promesse, sebbene anch’essi abbiano direzione di adattamento mondo-a-mente. È giusto – fino a un certo punto – per un parlante dire in maniera riflessiva “vorrei sia che tu rimanessi sia che te ne andassi”. Ma è irrazionale se egli dice simultaneamente “rimani!” e “vai via!”, e si è ugualmente irrazionali se si formano le intenzioni simultanee di andare e di restare, o se si fa la promessa simultanea di andare e di restare. Non si può

coerentemente avere delle intenzioni incoerenti o fare delle promesse incoerenti, emettere degli ordini incoerenti poiché le intenzioni, gli ordini e le promesse sono progettati per causare le azioni e non ci possono essere azioni incoerenti. Per la stessa ragione le intenzioni, gli ordini e le promesse impegnano l'agente nei confronti della credenza che l'azione è possibile ma che non è possibile effettuare entrambe le due azioni inconsistenti. I desideri e gli obblighi, in genere, non hanno tali condizioni. Si possono sostenere desideri inconsistenti ed avere degli obblighi inconsistenti.

Questa caratteristica ci fornisce la possibilità di un principio di distacco per le intenzioni? Se io intendo che p e credo che se p allora q , sono vincolato a intendere che q ? Io penso di no, tuttavia, la domanda è più complicata di quanto non possa sembrare a prima vista, e poiché si collega al famoso principio di Kant, mi soffermerò su una discussione su Kant.

VI. “Colui che vuole il fine vuole i mezzi”

Nessuna discussione della logica della ragione pratica sarebbe completa senza almeno un accenno al famoso principio di Kant secondo il quale colui che vuole il fine vuole i mezzi. Questo ci fornisce un principio logico deduttivo per la ragione pratica? Vale a dire, la frase “io voglio il fine F ” mi impegna logicamente nei confronti di “io voglio il mezzo M ” per lo meno nei casi in cui M è una condizione necessaria per ottenere F ? È analogo al modo in cui “io credo che p ” mi impegna nei confronti di “io credo che q ”, nei casi in cui q è una conseguenza logica di p ?

Ebbene tutto dipende da che cosa intendiamo per “voglio”. In base a un'interpretazione perfettamente naturale la dottrina è semplicemente falsa, per le ragioni che ho esposto in precedenza. Se volere è una questione che riguarda l'avere un desiderio fortissimo o una pro-attitudine verso qualche corso di azione futuro nel quale sono in grado di impegnarmi, allora semplicemente non è che quando voglio il fine sono logicamente vincolato a volere il mezzo. Come ho suggerito in precedenza, potrebbe essere che il mezzo non sia disponibile, o praticabile, per una qualche ragione. Desidero tantissimo eliminare i sintomi della mia influenza ma il solo modo per farlo è di commettere un suicidio, non essendoci una cura conosciuta, ma allo stesso tempo, non sono vincolato a volere il suicidio.

Se interpretiamo “voglio” come desiderio, il principio di Kant risulta dunque essere falso. Ma supponiamo di interpretarlo come un'intenzione, sia un'intenzione precedente che un'intenzione-in-azione. Supponiamo che io abbia un'intenzione precedente di fare F e che io creda che fare M sia una condizione necessaria per fare F . Sono vincolato nei confronti dell'intenzione di fare M ? Mi sembra che dobbiamo distinguere tra l'avere un vincolo a fare qualcosa che so che implicherà il fare M e avere un vincolo a fare M intenzionalmente. Banalmente, questo deriva dal fatto che io intendo fare F e so che il fare F implica necessariamente il fare M , e che ho un impegno a fare intenzionalmente qualcosa che implicherà M . Ma non per questo è necessario che io abbia un impegno a fare M intenzionalmente. Consideriamo allora l'esempio precedente della mia intenzione di fissare il vostro dente. Abbiamo le seguenti premesse:

Intendo (io fisso il vostro dente)

Cred (se fisso il vostro dente vi causerò dolore)

Ma non per questo sono vincolato alla conclusione

Intendo (io causo il vostro dolore)

Un'intenzione mi vincola nei confronti di un corso d'azione, ma non mi vincola a fare tutte le cose che so essere implicate nell'esecuzione dell'intenzione originaria. Così il fatto di avere l'intenzione di compiere p e di avere la credenza che se p allora q non mi vincola ad avere l'intenzione di compiere q . L'argomentazione per questa tesi, usando l'esempio esposto sopra, è che quando vi causo dolore, io non lo faccio intenzionalmente, ma solamente come sottoprodotto della mia intenzione reale. E l'argomentazione per quel punto, a sua volta, è che causarvi dolore non è parte delle condizioni di soddisfazione della mia intenzione, né è implicato dalle condizioni di soddisfazione della mia intenzione, perché se io fallisco nel causarvi dolore, non fallisco in ciò che stavo cercando di fare. Quando fisso il vostro dente io posso avere una ferma credenza il cui contenuto è che fissare il vostro dente vi causerà dolore, ma io non sono per questo impegnato nei confronti dell'intenzione di causarvi dolore. E la prova conclusiva la abbiamo se ci chiediamo qual è il successo o il fallimento della mia azione. Se fallisco nel causarvi dolore, non è la mia

intenzione originaria che ha fallito; ma piuttosto una delle mie credenze si è rivelata falsa. Semplicemente non accade, in generale, che chiunque voglia il fine (nel senso di avere un'intenzione di raggiungere quel fine) debba per questo volere tutto ciò che accade come parte conosciuta della realizzazione di quell'intenzione.

C'è tuttavia un tipo di caso in cui il principio di Kant è vero. Supponiamo che io abbia l'intenzione-in-azione di fissarvi il dente, e supponiamo che io abbia anche la credenza che la condizione necessaria per fissarvi il dente sia che io trapani *intenzionalmente* il vostro dente. Questo caso differisce da quello precedente perché il trapanare il vostro dente non è una parte collaterale del fissare il vostro dente così come lo era il causare il vostro dolore. Piuttosto, è un *mezzo* che deve essere intenzionato affinché venga realizzata l'intenzione originaria. Così, c'è un'interpretazione naturale del principio di Kant in cui esso si rivela corretto, e questa interpretazione è la seguente:

Se io considero un fine F e so che per raggiungere F devo intenzionalmente fare M , allora io sono impegnato a voler fare M . In questo senso mi sembra che “chi vuole il fine” sia vincolato a volere il mezzo.

VII. Conclusioni

La morale di questa discussione può essere esposta piuttosto brevemente. La logica deduttiva ha a che fare con le relazioni logiche tra proposizioni, predicati, insiemi, ecc. In senso stretto non esiste una logica deduttiva della ragione pratica, ma allora, in senso stretto, non esiste nemmeno una logica deduttiva della ragione teorica. A causa della combinazione del vincolo e della direzione di adattamento delle credenze, è tuttavia possibile tracciare una mappa delle relazioni logiche che accadono nella ragione teorica tramite la logica deduttiva e questo risulta invece impossibile per la ragione pratica. Perché c'è questa differenza? Il desiderio è differente dalla credenza per due importanti aspetti. Il desiderio ha una direzione di adattamento verso l'alto, e una persona con un desiderio non è impegnata nei confronti della soddisfazione di quel desiderio nello stesso modo in cui una persona che possiede una credenza è impegnata nei confronti della verità di quella stessa credenza. Questo tiene conto delle due caratteristiche del desiderio che abbiamo notato in precedenza, la necessità

dell'inconsistenza, o incoerenza, e la non distaccabilità del desiderio. Le intenzioni assomigliano un po' di più ai desideri perché non implicano un vincolo nei confronti della soddisfazione dell'intenzione. Nondimeno, la persona che ha un'intenzione non è impegnata nei confronti dell'intenzione di realizzare tutte le conseguenze della realizzazione della sua intenzione. Egli è impegnato soltanto nei confronti dei mezzi che sono necessari per la realizzazione dei suoi scopi. Per queste ragioni non ci sarà una "logica deduttiva della ragione pratica" nemmeno nel senso limitato entro il quale è possibile avere una logica deduttiva della ragione teorica.

P 47 9.

La coscienza, la libera azione e il cervello

I. La coscienza e il cervello

Buona parte di questo libro riguarda il divario (*gap*) e le sue implicazioni per lo studio della razionalità. Poiché il divario è una caratteristica della coscienza umana, in questo senso il libro riguarda anche la coscienza. Il divario è quella caratteristica della coscienza delle azioni volontarie, tramite cui le azioni sono esperite come non aventi condizioni causali psicologiche sufficienti a determinarle. Questo è parte di ciò che si intende quando si afferma che esse sono, almeno psicologicamente, libere. Non c'è dubbio che il divario sia psicologicamente reale, ma lo è anche empiricamente? Esso è anche una realtà neurobiologica oltre che psicologica? Se la libertà umana esiste realmente, essa deve essere una caratteristica delle funzioni del cervello. Lo scopo di questo capitolo è quello di collocare una spiegazione della coscienza volontaria, o della coscienza della libera azione, all'interno di una spiegazione generale della coscienza e, a sua volta, di inserire tale spiegazione all'interno di una spiegazione delle funzioni cerebrali.

Poiché stiamo per intraprendere la discussione di un problema filosofico tradizionale, è una buona idea fare un piccolo passo indietro e chiederci perché abbiamo ancora un tale problema. Nel capitolo 1 ho sostenuto che tali problemi, solitamente, sorgono quando abbiamo un conflitto tra due teorie apparentemente inconsistenti, e sentiamo di non poter rinunciare a nessuna delle due. In questo caso la credenza nel libero arbitrio è basata sulle nostre esperienze coscienti del divario, ma noi disponiamo anche di una assunzione metafisica fondamentale secondo la quale l'universo è un sistema fisico chiuso, interamente determinato dalle leggi della fisica. Che fare? La prima cosa da notare è che le leggi più fondamentali della fisica, al livello della meccanica quantistica, non sono deterministiche. La seconda cosa da notare è che le leggi della fisica, in realtà, non determinano nulla. Le leggi sono un insieme di enunciati che descrivono le relazioni tra varie quantità fisiche, e

talvolta questi enunciati descrivono condizioni causalmente sufficienti in situazioni particolari, talvolta no. La terza cosa da notare è che la tesi secondo la quale l'universo è un sistema fisico chiuso, per quanto abbia un significato chiaro, è una proposizione che, attraverso i secoli, abbiamo reso vera per convenzione. Non appena pensiamo che qualcosa esista realmente nel mondo empirico e pensiamo di comprenderlo, anche se solo lontanamente, lo chiamiamo "fisico". In quanto parti del mondo reale, la coscienza, l'intenzionalità e la razionalità sono fenomeni "fisici", come qualsiasi altra cosa. Queste riflessioni non risolvono il nostro problema ma dovrebbero portarci a pensarlo in un senso meno ristretto. Iniziamo chiedendoci in che modo la coscienza si inserisce nell'universo "fisico".

Negli ultimi dieci anni circa, è emersa una particolare concezione della coscienza e delle sue relazioni con il cervello ed essa è stata sempre più comunemente accettata sia in filosofia che nelle neuroscienze. Essa è profondamente opposta sia al dualismo che al materialismo, così come essi sono stati tradizionalmente concepiti. In particolare, essa si oppone a quelle concezioni della coscienza che tentano di negare l'irriducibile soggettività degli stati coscienti, o che tentano di ridurre la coscienza al comportamento, a un programma per computer o a stati funzionali di un sistema. Questa concezione della coscienza sta diventando sempre più comunemente accettata, ma è ancora controversa.

Essa è la seguente: la coscienza è un fenomeno biologico reale. Consiste di stati interni, qualitativi, soggettivi, stati unificati di sensibilità, consapevolezza, pensieri e sentimenti. Questi stati iniziano quando ci svegliamo al mattino da un sonno privo di sogni e continuano attraverso tutto il giorno, fino a che non ritorniamo nuovamente incoscienti. Secondo questa spiegazione i sogni sono una forma di coscienza, sebbene essi siano, sotto diversi aspetti, differenti dalla normale coscienza della veglia. Le caratteristiche chiave della coscienza, in base a questa concezione, sono che essa è qualitativa, soggettiva e unificata nel senso che spiegherò ora. Per ogni stato cosciente c'è un certo sentimento qualitativo per quello stato. C'è qualcosa che è come, per cui ci si sente come, essere in uno stato di quel tipo. Questo è vero per i pensieri, come il pensiero che due più due fa quattro, così come per il gusto della birra, il profumo della rosa o la visione del cielo blu. Tutti gli stati coscienti, siano essi percezioni o processi di pensiero, sono, nel senso che sto cercando di spiegare, stati qualitativi. Essi sono inoltre soggettivi, nel senso che esistono solamente in quanto esperiti da un soggetto umano o animale. Essi hanno una caratteristica aggiuntiva che vale la

pena enfatizzare: le esperienze coscienti, come il gusto della birra o il profumo della rosa, si presentano sempre come parte di un campo cosciente unificato. Io, per esempio, non sento proprio ora la pressione della camicia sulla mia schiena, il gusto di caffè in bocca e la visione dello schermo del computer di fronte a me; io ho tutte queste cose come parti di un singolo campo cosciente unificato.

Qual è la relazione tra la coscienza, definita in questo modo, e i processi cerebrali? Riconoscerete in questo il tradizionale problema mente-corpo. Credo che nella sua forma filosofica (sebbene – ahimè – non nella sua forma neurobiologica), il problema mente-corpo abbia una soluzione piuttosto semplice. Ed è la seguente: tutti i nostri stati coscienti sono causati da processi neuronali di livello inferiore del cervello, e sono essi stessi delle caratteristiche del cervello. E' possibile osservare questo in maniera quasi ovvia nel caso dei dolori. I mie dolori attuali sono causati da una serie di scariche neuronali che cominciano dalle terminazioni nervose periferiche e continuano fino alla colonna spinale, attraverso il tratto di Lissauer, fino al talamo e altre regioni basali del cervello. Alcuni di questi si estendono alla corteccia sensoria, specialmente la Zona 1, e questa sequenza causa il fatto che io provo dolore. Che cosa sono questi dolori? I dolori stessi sono semplicemente delle caratteristiche del cervello di livello superiore o di sistema. Le esperienze soggettive, qualitative del dolore nel campo cosciente totale sono causate da processi neurobiologici del cervello e dal resto del sistema nervoso centrale e sono essi stessi, in quanto elementi del campo unificato della coscienza, delle caratteristiche del sistema di neuroni e di altre cellule che costituiscono il cervello umano.

Che cosa sono, esattamente, i processi neuronali che causano queste esperienze coscienti? Attualmente non abbiamo la risposta a questa domanda. Stiamo facendo dei progressi, ma fino ad oggi i progressi sono stati lenti. Attualmente ci sono, per quanto io ne sappia, almeno due approcci generali al problema della coscienza, e per poter entrare nella parte principale di questa discussione, devo dire qualcosa di più riguardo ad essi. Chiamerò il primo approccio "l'approccio a mattoni" (*Building block*). L'idea è che il nostro campo cosciente consiste di una serie di componenti separati, che sono le esperienze coscienti individuali. Questi elementi costituiscono il campo totale nello stesso modo in cui i mattoni di una casa costituiscono la casa. L'assunzione che sta alla base del progetto di ricerca "a mattoni" è che se noi potessimo trovare esattamente il modo in cui funziona anche un solo mattone, o blocco, per esempio, come esperiamo visivamente il colore rosso, questo

potrebbe darci la chiave per risolvere l'intero problema della coscienza, poiché i meccanismi per mezzo dei quali vengono prodotte le esperienze coscienti del rosso assomiglierebbero presumibilmente ai meccanismi per mezzo dei quali vengono prodotte le esperienze dei suoni o dei gusti. L'idea è di trovare il correlato neuronale della coscienza (CNC) per le esperienze sensoriali individuali, e poi operare una generalizzazione da queste fino alla coscienza in senso generale.

Per ragioni che ho cercato di spiegare altrove¹ penso che l'approccio "a mattoni", o a blocchi separati, sia sbagliato. Ogni mattone è presente solo in un soggetto che è già cosciente. Non credo che possiamo scoprire i meccanismi che producono la coscienza cercando, ad esempio, di scoprire i meccanismi che producono l'esperienza del rosso, poiché solamente un soggetto che è *già* cosciente può avere l'esperienza del rosso. L'approccio a mattoni vorrebbe sostenere che in un soggetto non cosciente, se si potesse produrre il CNC per un singolo mattone, diciamo l'esperienza del rosso, allora il soggetto avrebbe improvvisamente un lampo di rosso e nessun altro stato cosciente. Questa è un'ipotesi empirica possibile, ma mi sembra per lo più improbabile, dato il nostro grado di conoscenza sul funzionamento del cervello. Mi sembra molto più attendibile l'idea che arriveremo a comprendere in che modo il cervello causa la coscienza se potremo trovare la differenza tra il comportamento neurofisiologico di un cervello non cosciente e quello di un cervello cosciente. Ciò che ci piacerebbe davvero conoscere è, in primo luogo, in che modo il soggetto diviene cosciente. Una volta che il soggetto è cosciente, possono essere indotte quelle esperienze particolari che modificheranno il campo cosciente unificato esistente.

C'è un'altra linea di ricerca che chiamerò l'"approccio del campo unificato". Invece di pensare alla coscienza come costituita da una serie di mattoncini, una costruzione di mattoni o di blocchi, dovremmo considerare seriamente l'unità di cui parlavo prima e pensare all'intero campo cosciente come a un'unità. Dovremmo pensare agli stimoli percettivi individuali non in quanto *creanti* la coscienza, ma in quanto *modificanti* una coscienza pre-esistente. Secondo questa spiegazione, invece di cercare il CNC del rosso, per esempio, dovremmo provare a trovare la differenza tra un cervello cosciente e uno non cosciente.

Secondo la spiegazione che sto presentando, le tre caratteristiche che ho menzionato, la qualitatività, la soggettività e l'unità, non sono tre caratteristiche distinte, bensì differenti

¹ "Consciousness", *Annual Review of Neuroscience* 2000, vol. 23, pp. 557-578.

aspetti della medesima caratteristica. Una volta che una caratteristica è qualitativa, nel senso che ho spiegato, essa deve essere anche soggettiva, perché la nozione di qualitatività di cui stiamo parlando è qualcosa che viene esperito da un soggetto. E una volta che ci sono esperienze che sono sia qualitative che soggettive esse sono necessariamente unificate. Si può osservare questo ancora una volta grazie ad un esperimento mentale. Se immaginate il vostro presente stato di coscienza come spezzettato in diciassette parti, voi immaginerete diciassette differenti campi coscienti. La qualitatività, la soggettività e l'unità, non sono caratteristiche differenti; sono piuttosto aspetti di una sola caratteristica, e questa caratteristica è la reale essenza della coscienza.

II. La coscienza e l'azione volontaria

Quando esploriamo la caratteristica del campo cosciente, scopriamo un fatto di particolare rilievo. C'è una differenza sorprendente e drammatica tra la caratteristica qualitativa delle esperienze percettive e la caratteristica qualitativa delle azioni volontarie. Nel caso delle esperienze percettive, io sono un recipiente passivo delle esperienze che vengono causate dall'ambiente esterno. Così, se io tengo una mano davanti alla faccia, non dipende da me se io vedo o meno quella mano. L'apparato percettivo e gli stimoli esterni sono sufficienti di per sé a causare in me un'esperienza visiva della mia mano davanti alla faccia. Non ho a disposizione una scelta; le cause sono sufficienti a produrre l'esperienza.

Se, d'altro canto, decido di sollevare il mio braccio destro sopra la mia testa, questo dipende interamente da me. Dipende da me il fatto che io sollevi il mio braccio destro o quello sinistro, quanto li sollevo ecc. L'azione volontaria, semplicemente, ha un differente campo cosciente rispetto a quello della percezione. Non sto, certamente, suggerendo che non esiste nessun elemento volontaristico della percezione. Penso che ci sia. Per esempio, negli esempi mutevoli della *Gestalt* si può, a piacere, cambiare la propria percezione da quella di un'anatra a quella di un coniglio e il contrario. In questo momento voglio semplicemente richiamare l'attenzione su alcune sorprendenti caratteristiche dell'azione volontaria che si trovano in netto contrasto con l'esperienza della percezione.

Il divario di cui abbiamo discusso sorge solamente con l'azione volontaria. Per prima cosa c'è un divario tra le ragioni per una decisione e la decisione; in secondo luogo c'è un

divario tra la decisione e la sua esecuzione; e, terzo, c'è un divario tra il cominciamento di un'azione e la sua continuazione fino al completamento. In fondo penso che i tre divari siano manifestazioni dello stesso fenomeno, poiché sono tutte e tre le manifestazioni della coscienza volitiva.

Come abbiamo visto nel capitolo 3, la struttura logica della spiegazione del comportamento umano, in cui l'agente agisce volontariamente sulla base di una ragione, ci richiede di postulare un sé irriducibile. Ora noi possiamo aggiungere a questa nozione puramente formale di sé il fatto che il sé, concepito in questo modo, richiede il campo unificato della coscienza. Avevamo postulato un sé per rendere intelligibile il fenomeno delle libere azioni razionali. Ma il sé postulato in questo modo richiede un campo cosciente unificato. Il sé non è identico al campo della coscienza, ma le sue operazioni, per mezzo delle quali si prendono decisioni sulla base delle ragioni e si agisce per realizzare quelle decisioni, richiede un campo unificato contenente sia elementi cognitivi come le percezioni e le memorie, sia elementi volitivi, come le decisioni e le azioni. Perché? Ebbene, se provate a immaginare la mente come un fascio humaneo di percezioni sconnesse, non c'è alcun modo in cui il sé possa operare nel fascio. Affinché il sé possa operare nel prendere decisioni, dovrete avere un sé differente per ogni elemento differente del fascio.

III. Il libero arbitrio

Vorrei ora applicare le lezioni finora apprese alla discussione del tradizionale problema della libertà della volontà. Ci sono, senza dubbio, molti sensi differenti di "libero arbitrio" e di "determinismo"; ma in questa discussione, il problema della libertà della volontà sorge per quelle parti del campo cosciente in cui sperimentiamo il divario. Questi sono i casi che sono tradizionalmente chiamati "volizione". Non c'è dubbio che noi abbiamo esperienze del tipo di quelle che ho chiamato le esperienze del divario; cioè, noi esperiamo le nostre normali azioni volontarie nello stesso modo in cui avvertiamo la possibilità di azioni alternative davanti a noi e avvertiamo che gli antecedenti psicologici dell'azione non sono sufficienti a determinare l'azione. Si noti che sulla base di questa spiegazione il problema del libero arbitrio sorge solamente per la coscienza, e sorge solamente per la coscienza volitiva o attiva; esso non sorge per la coscienza percettiva.

Che cos'è, allora, esattamente il problema della libertà della volontà? Normalmente si ritiene che il libero arbitrio sia opposto al determinismo. La tesi del determinismo circa le azioni è che ogni azione è determinata da condizioni causali antecedenti sufficienti. Per ogni azione, le condizioni causali dell'azione in quel contesto sono sufficienti a produrre quell'azione. Così, per quanto riguarda le azioni, niente potrebbe accadere in maniera differente da come di fatto accade. La tesi del libero arbitrio, a volte chiamata "libertarismo", sostiene che alcune azioni, per lo meno, sono tali che le condizioni causali antecedenti dell'azione non sono causalmente sufficienti a produrre l'azione. Riconosciuto che l'azione è accaduta, ed è accaduta per una ragione, l'agente avrebbe potuto tuttavia fare qualcos'altro, dati gli stessi antecedenti causali dell'azione.

La teoria contemporanea maggiormente diffusa riguardo al tema del libero arbitrio, è chiamata "compatibilismo". La tesi del compatibilismo sostiene che, se comprendiamo propriamente questi termini, la libertà della volontà è completamente compatibile con il determinismo. Dire che un'azione è determinata è semplicemente dire che ha delle cause come qualsiasi altro evento, e dire che è libera è semplicemente dire che è determinata da certi tipi di cause e non da altre. Così, se qualcuno mi punta una pistola alla testa e mi dice di alzare il mio braccio, la mia azione non è libera; ma se io sollevo il mio braccio per votare, come si dice, "liberamente", o "per mia libera volontà", allora la mia azione è libera. Sebbene in entrambi i casi, sia nel caso del votare, che in quello della pistola puntata alla testa, la mia azione sia completamente determinata causalmente.

Penso che il compatibilismo semplicemente non colga il punto essenziale del problema del libero arbitrio. Così come l'ho definito, il libertarismo è totalmente inconsistente con il determinismo. Per ripeterlo, il determinista sostiene che "ogni azione è preceduta da condizioni causalmente sufficienti che determinano l'azione". E il libertarista asserisce la negazione di ciò: "per alcune azioni, le condizioni causali antecedenti non sono sufficienti a determinare l'azione".

Penso che senza dubbio vi sia un senso di "libero" e di "determinato" in cui il compatibilismo è giusto. Quando per esempio la gente marcia per le strade sventolando degli striscioni e chiedendo "libertà", questa gente solitamente non è molto interessata alle leggi della fisica. Essi abitualmente vogliono meno restrizioni governative sulle loro azioni, o qualcosa del genere; e non sono preoccupati degli antecedenti causali delle loro azioni. Ma questo senso di "libertà", quando significa assenza di vincoli esterni, è irrilevante per il

problema del libero arbitrio, così come io l'ho enunciato. Non riesco a pensare a nessun interessante problema filosofico del libero arbitrio al quale il compatibilismo possa fornire una risposta sostanziale.

Siamo giunti alla convinzione che esista la libertà della volontà, intesa nel mio senso, a causa dell'esperienza del divario. Così, il problema della libertà della volontà può essere posto come segue: quale realtà corrisponde a quelle esperienze? Dato che abbiamo esperienza delle nostre azioni come azioni prive di condizioni causali psicologiche antecedentemente sufficienti, perché dovremmo prendere seriamente questo fatto psicologico? Non è possibile che i fondamenti neurobiologici della psicologia siano causalmente sufficienti per determinare l'azione, anche se il livello psicologico di per sé non è causalmente sufficiente? E non potrebbero esserci delle cause psicologiche non coscienti che determinano l'azione? Anche se concediamo la realtà psicologica del divario, il problema del libero arbitrio rimane irrisolto. Che cos'è esattamente e come, esattamente, potremmo riuscire a risolverlo?

Per rendere il problema completamente chiaro, consideriamo il seguente esempio. Supponiamo che mi venga offerta la scelta al tempo t_1 tra due bicchieri di vino, un Borgogna e un Bordeaux, che stanno sul tavolo di fronte a me. Supponiamo che io li trovi entrambi attraenti e che, dopo 10 secondi, al tempo t_2 , io decida per il Borgogna, vado verso il tavolo, sollevo il bicchiere e lo bevo. Chiamiamo questa l'azione A, e supporremo che essa inizi a t_2 e continui per pochi secondi fino a t_3 . Per semplicità supporremo che non ci sia alcun divario temporale psicologico tra la decisione e la sua esecuzione. Nell'istante in cui ho deciso per il Borgogna, a t_2 , è iniziata l'intenzione-in-azione ed io ho preso il bicchiere. (Nel tempo reale, naturalmente, c'è un divario temporale di circa 200 millisecondi tra l'inizio della mia intenzione-in-azione e l'effettivo inizio dei movimenti muscolari). Supponiamo anche che questa fosse un'azione volontaria con un divario: io non ero sotto la costrizione di un'ossessione o di altre cause sufficienti a determinare l'azione. Noi, semplicemente, stabiliremo che in questo esempio non c'erano delle cause psicologiche non coscienti sufficienti a determinare l'azione. La mia azione era libera nel senso che le cause psicologiche, coscienti e non coscienti, operanti su di me, non erano sufficienti a determinare l'azione A. Che cosa significa questo, esattamente? Almeno quanto segue. Una completa specificazione di tutte le cause psicologiche operanti su di me a t_1 , con tutti i loro poteri causali, inclusa qualsiasi legge psicologica rilevante per il caso, non sarebbe

sufficiente a implicare il fatto che io avrei compiuto l'azione *A sotto qualsiasi descrizione*. Non riuscirebbero a implicare non solo “JRS sceglierà il Borgogna”, ma nemmeno “questo braccio si muoverà in questa direzione e queste dita si chiuderanno intorno a questo oggetto”. Secondo questo aspetto, le cause psicologiche a t_1 sono diverse dalle usuali cause fisiche. Se nel prendere il Borgogna urtassi inavvertitamente un bicchiere vuoto facendolo cadere dal tavolo, la descrizione delle cause operanti sul bicchiere che iniziasse al momento dell'impatto sarebbe sufficiente a implicare il fatto che il bicchiere cadrà a terra.

Ho detto prima che tutti questi processi psicologici sono essi stessi causati dal cervello e realizzati in esso. Così, al momento t_1 , la mia percezione cosciente dei due bicchieri di vino rosso e le mie riflessioni coscienti sul loro rispettivo valore, sono state causate da processi neurobiologici di livello inferiore del cervello e realizzati nella struttura del cervello. Ed ecco ora il problema: assumendo che non ci fossero ulteriori stimoli per il cervello, ad esempio ulteriori percezioni, i processi neurobiologici che accadevano in me a t_1 erano causalmente sufficienti a determinare lo stato totale del mio cervello a t_2 ? E lo stato totale del mio cervello a t_2 era sufficiente a causare la continuazione dei processi cerebrali che sono accaduti tra t_2 e t_3 ? Se le cose stanno in questo modo, c'è una descrizione dell'atto *A* sotto la quale esso ha condizioni causali sufficienti antecedenti: lo stato del mio cervello a t_2 era tale che i neurotrasmettitori causavano l'inizio delle contrazioni muscolari che costituivano i componenti corporei dell'atto *A*, e la continuazione dei processi da t_2 a t_3 era sufficiente a causare la continuazione delle contrazioni muscolari fino al completamento dell'azione a t_3 . Il problema della libertà della volontà si riduce a questo alla seguente domanda: assumendo che nessun ulteriore stimolo esterno rilevante arrivi al cervello, lo stato del cervello a t_1 , descritto in termini neurobiologici, era causalmente sufficiente a determinare lo stato del cervello a t_2 ? E tale stato a t_2 era sufficiente a portarlo a t_3 ? *Se la risposta a queste domande è sì, per questo e per altri casi simili rilevanti, allora non abbiamo il libero arbitrio*. Il divario reale dal punto di vista psicologico non corrisponde a nessuna realtà neurobiologica e la libertà della volontà è un'illusione di massa. *Se la risposta a quella domanda è no, allora, date certe assunzioni circa il ruolo della coscienza, noi abbiamo davvero il libero arbitrio*.

Ora, perché ogni cosa dovrebbe ridursi a questo? Perché lo stato del cervello a t_2 era sufficiente a causare i movimenti muscolari che danno inizio all'azione, e gli stati del cervello da t_2 a t_3 sono sufficienti a far sì che i movimenti muscolari portino al

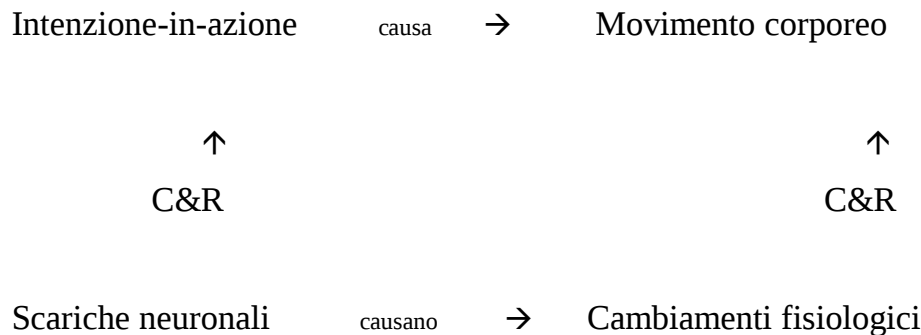
completamento dell'azione. Una volta che l'acetilcolina colpisce i terminali assonali dei neuroni motori, allora, assumendo che il resto dell'apparato psicologico funzioni normalmente, i muscoli si muoveranno per diretta necessità causale. I primi due divari accadono prima dell'inizio dei movimenti muscolari, e il terzo divario ha luogo tra l'inizio dell'azione e la sua continuazione fino al completamento. Il divario è un fenomeno psicologico reale, ma se è un fenomeno reale che ha un effetto sul mondo, deve avere un correlato neurobiologico. Come domanda neurobiologica, la realtà del divario si riduce a questo: gli stati del cervello da t_1 a t_3 sono sufficienti a far sì che ognuno di essi determini lo stato successivo tramite condizioni causalmente sufficienti? Il problema della libertà della volontà, in neurobiologia, è un problema che riguarda direttamente le relazioni tra certi tipi di coscienza e i processi neurobiologici. Se questa è una domanda interessante, essa è una domanda scientifica che riguarda la causazione di certi tipi di azioni coscienti. Intendo dunque ora affrontare questo argomento con molta attenzione e cercare di giungerne ad una comprensione essenziale.

IV . Ipotesi 1: il libertarismo psicologico con il determinismo neurobiologico

Per cominciare, dobbiamo ricordarci di quanto abbiamo appreso fino ad ora. Tutti i nostri stati di coscienza sono causati da processi neurobiologici di livello inferiore e superiore del cervello. Essi stessi possono causare stati di coscienza conseguenti o movimenti corporei perché essi sono fondati nella neurobiologia. Così, nei casi in cui non ci sono divari, la causazione sinistra-destra attraverso il tempo a livello superiore combacia esattamente con la causazione sinistra-destra attraverso il tempo a livello inferiore. Per esempio, la mia intenzione-in-azione è causata da processi del cervello di livello inferiore. Essa, a sua volta, causa il fatto che il mio braccio si sollevi. I processi neurobiologici che causano l'intenzione-in-azione a loro volta causano una serie di cambiamenti fisiologici che causano e realizzano i movimenti del mio braccio. Queste relazioni sono tipiche di qualsiasi sistema che abbia livelli di descrizione causalmente reali. Così, un motore d'automobile ha lo stesso insieme di relazioni formali. Nessun epifenomeno è il risultato di queste relazioni. L'intenzione-in-azione è causalmente reale quanto lo è la solidità del pistone. Inoltre non c'è alcuna sovradeterminazione causale. Non stiamo parlando di sequenze causali indipendenti, ma piuttosto delle stesse sequenze causali descritte a livelli differenti. Ancora una volta,

l'analogia del motore dell'automobile funziona perfettamente. Possiamo descrivere la causazione al livello delle molecole o possiamo descriverla al livello dei pistoni e dei cilindri. Queste non sono sequenze causali indipendenti, ma la stessa sequenza causale descritta a livelli differenti.

In scritti precedenti² ho rappresentato queste relazioni che costituiscono l'azione volontaria come un parallelogramma che appare in questo modo:

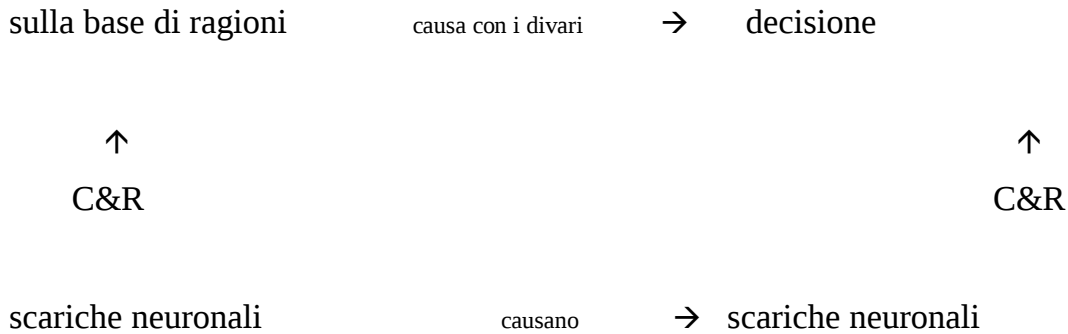


Al livello superiore, l'intenzione-in-azione causa il movimento corporeo, al livello inferiore le scariche neuronali causano i cambiamenti fisiologici, e ad ogni punto il livello inferiore causa e realizza (C&R) il livello superiore. Così rappresentata, l'intera struttura è deterministica ad ogni livello.

Che dire dei casi con un divario, quando delibero e quindi prendo una decisione, per esempio? Mi sembra che ci siano almeno due possibilità. La prima possibilità (ipotesi 1) è questa: all'indeterminatezza del livello psicologico corrisponde un sistema completamente deterministico a livello neurobiologico. Così, sebbene abbiamo un divario psicologico tra le ragioni per l'azione e la decisione, noi non abbiamo alcun divario al livello neurobiologico tra la realizzazione neurofisiologica delle ragioni per l'azione sottoforma di credenze e desideri e la successiva realizzazione neurofisiologica della decisione. Ecco come apparirebbe:

deliberazione

² John R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, New York, Cambridge University Press, 1983, p. 270, tr. it. *Della intenzionalità: un saggio di filosofia della mente*, Bompiani, Milano 1985, pp. 270-271.



In questo caso il divario produce un'asimmetria tra il parallelogramma dell'azione volontaria e il parallelogramma della cognizione. Si può notare questo se si mettono in contrasto la decisione e l'azione con la memoria. Supponiamo che io veda una scena drammatica, diciamo un incidente d'auto, e che, successivamente, io abbia un ricordo dell'incidente d'auto che ho visto. Ho avuto un'esperienza psicologica, l'esperienza percettiva, e tale evento psicologico ha costituito il fondamento causale sufficiente per l'evento psicologico seguente del ricordo dell'incidente osservato. Ma noi sappiamo che tutto questo è possibile perché abbiamo una sequenza di condizioni causalmente sufficienti nella neurobiologia. La percezione reale, in termini neurobiologici, è sufficiente a lasciare delle tracce di ricordo nella memoria a breve termine e nella memoria a lungo termine, per mezzo delle quali io ricordo l'evento psicologico. Vale a dire, nel caso della cognizione, così come per la relazione della percezione e della memoria, le condizioni sufficienti a livello superiore, o psicologico, sono corrisposte da condizioni sufficienti a livello inferiore o neurofisiologico. Si ha un parallelogramma perfetto. Per la volizione, contrariamente alla cognizione, non si ha un tale parallelogramma. Nel caso della volizione, l'indeterminatezza psicologica coesisterebbe con il determinismo neurobiologico.

Se questo è il modo in cui la natura funziona, allora avremmo una specie di compatibilismo. Il libertarismo psicologico sarebbe compatibile con il determinismo neurobiologico. I processi psicologici, sebbene siano essi stessi causati da processi neuronali di livello inferiore, non sarebbero, ciononostante, condizioni causali sufficienti per il conseguente evento psicologico dell'azione intenzionale. A t_1 i processi psicologici tramite i quali sto decidendo quale bicchiere di vino bere sono determinati in maniera causale interamente da processi neuronali di livello inferiore per mezzo della causazione inferiore-superiore. A t_2 io decido per il Borgogna. Questa decisione, di nuovo, è interamente fissata

dalla causazione inferiore-superiore, anche se c'è un divario a livello psicologico tra la mia riflessione sulle ragioni e la mia decisione. Da t_2 a t_3 i componenti del movimento muscolare dell'azione A, ovvero il mio prendere il vino che ho in mano e portarlo alla bocca, sono causati da processi neurobiologici, tramite la causazione inferiore-superiore, anche se, di nuovo, c'è un divario a livello psicologico tra l'inizio dell'azione e la sua continuazione fino al completamento. Così abbiamo dei divari al livello psicologico, ma nessun divario sotto forma di causazione inferiore-superiore tra il livello neurobiologico e il livello psicologico, e nessun divario a livello neurobiologico tra qualsiasi stato del sistema e lo stato successivo del sistema stesso. Questo ci darebbe il determinismo fisiologico con il libertarismo psicologico.³

Questo risultato, comunque, è intellettualmente molto insoddisfacente, poiché, detto in una parola, esso è una forma modificata di epifenomenismo. Ci dice che i processi psicologici del prendere decisioni razionali non hanno grande importanza, che l'intero sistema è deterministico a livello inferiore, e l'idea che il livello superiore abbia un elemento di libertà è semplicemente un'illusione sistematica. Mi sembra che a t_1 io abbia a disposizione una scelta tra il Borgogna e il Bordeaux e che le cause operanti su di me non siano sufficienti a determinare ogni movimento corporeo così come un qualsiasi processo di pensiero da t_1 a t_2 a t_3 . Se l'ipotesi 1 è vera, allora ogni movimento muscolare, così come ogni pensiero cosciente, inclusa l'esperienza cosciente del divario, l'esperienza della "libera" scelta decisionale, è interamente fissato in anticipo; e la sola cosa che possiamo dire circa l'indeterminismo psicologico a livello superiore è che esso ci dà un'illusione sistematica del libero arbitrio. La tesi è di tipo epifenomenico sotto questo aspetto: c'è una caratteristica delle nostre vite coscienti, nel prendere delle decisioni razionali e nel tentativo di realizzare tali decisioni, in cui abbiamo esperienza del divario ed abbiamo esperienza dei processi come se questi producessero una differenza causale per il nostro comportamento, ma, in realtà, essi non producono nessuna differenza. I movimenti corporei saranno esattamente gli stessi indipendentemente dal modo in cui questi processi hanno avuto luogo.

Probabilmente questo è ciò che potrebbe risultare, ma se fosse veramente così, l'ipotesi andrebbe contro tutto quello che conosciamo riguardo all'evoluzione. Avrebbe la conseguenza che l'incredibilmente elaborato, complesso, sensibile e – soprattutto –

³ Ho detto che questa è una forma di compatibilismo, ma essa differisce dal compatibilismo tradizionale perché la versione tradizionale postula il determinismo ad ogni livello. Questa versione postula invece l'indeterminismo psicologico e il determinismo neurobiologico.

biologicamente dispendioso sistema umano e animale del prendere decisioni razionali coscienti non produrrebbe effettivamente nessuna differenza nella vita e nella sopravvivenza degli organismi. L'epifenomenismo è una tesi possibile, ma è assolutamente incredibile e improbabile, e se l'accettassimo seriamente, essa produrrebbe un cambiamento nella nostra visione del mondo, e cioè nella nostra concezione delle nostre relazioni col mondo, più radicale di qualsiasi cambiamento precedente, inclusa la rivoluzione copernicana, la teoria della relatività einsteiniana e la meccanica quantistica.

Perché l'ipotesi 1 renderebbe la coscienza più epifenomenica di qualsiasi altra caratteristica di livello superiore di un sistema fisico? Dopo tutto, la solidità del pistone nel motore dell'automobile è interamente spiegata dal comportamento delle molecole ma questo non rende epifenomenica la solidità. La differenza è questa: le caratteristiche essenziali della solidità influiscono sulla prestazione del motore, ma la caratteristica essenziale del prendere delle decisioni coscienti, l'esperienza del divario, non influirebbe per nulla sulla prestazione dell'agente. I movimenti corporei sarebbero stati gli stessi indipendentemente dalle esperienze del divario.

V. Ipotesi 2: Un sistema di causazione con coscienza e indeterminatezza

Secondo la teoria alternativa (ipotesi 2), all'assenza di condizioni causalmente sufficienti a livello psicologico corrisponde una parallela mancanza di condizioni causalmente sufficienti a livello neurobiologico. Ma questo che cosa potrebbe mai voler dire? Come dovrebbe essere il diagramma di una tale ipotesi? A questo punto mi sembra che dobbiamo esaminare criticamente le assunzioni insite nella nostra rappresentazione grafica con le sue metafore di "inferiore-superiore", "dall'alto verso il basso", "livelli di descrizione", ecc. Penso che esse si riveleranno inadeguate a questo stadio dell'indagine. Il problema è questo: l'idea che la coscienza sia una caratteristica di livello superiore, o di superficie, del cervello ci fornisce un'immagine della coscienza analoga alla vernice sulla superficie del tavolo. Allora la questione della causazione dall'alto verso il basso e inferiore-superiore diverrebbe quella di raggiungere qualcosa a livello superiore o inferiore. Tutto questo è sbagliato. La coscienza non si trova sulla superficie del cervello più di quanto la liquidità non si trovi sulla superficie dell'acqua. Piuttosto l'idea che stiamo cercando di

esprimere è che la coscienza è una *caratteristica di sistema*. È una caratteristica dell'intero sistema ed è presente – letteralmente – in tutti i punti rilevanti del sistema allo stesso modo in cui l'acqua in un bicchiere è liquida in ogni parte. La coscienza non esiste in una singola sinapsi più di quanto la liquidità non esista in una singola molecola. Ma allora l'immagine dei differenti livelli che si muovono in parallelo, che è rappresentata dal nostro diagramma, è sbagliata. L'intero sistema si muove compatto. Ciò che dobbiamo supporre, se pensiamo che la nostra esperienza cosciente della libertà non sia una completa illusione, è che l'intero sistema si muove verso il prendere decisioni, e verso l'implementazione della decisione in azioni reali; dobbiamo supporre che la razionalità cosciente al livello superiore sia tuttavia realizzata ai livelli più bassi, e ciò significa che l'intero sistema si muove in un modo che è causale, ma non basato su condizioni causalmente sufficienti.

Per porre la questione di come il divario possa funzionare all'interno della neurobiologia, dobbiamo essere chiari riguardo a come esso funziona nella psicologia cosciente. Nel caso della razionalità cosciente non vi è nulla che colmi il divario. Una persona semplicemente prende una decisione e poi semplicemente agisce. Questi fatti ci risultano intelligibili solamente se postuliamo un agente razionale cosciente, capace di riflettere sulle proprie ragioni e poi di agire sulla base di queste ragioni. Sono riluttante ad usare la terminologia tradizionale, ma precedentemente ho sostenuto che quel postulato equivale al postulato di un sé. Noi possiamo comprendere l'azione cosciente razionale, libera, solo se postuliamo un sé cosciente. Ma questo postulato ha senso solo in relazione all'esistenza di un campo cosciente unificato di soggettività. Non si può spiegare il sé razionale solamente nei termini di un fascio humaneo di percezioni sconnesse. Così la seconda ipotesi non consiste nel fatto che vi è una spaccatura tra l'indeterminatezza al livello della psicologia e la determinatezza al livello della neurobiologia, ma piuttosto che l'intero sistema si muove compatto, e in quanto sistema cosciente e razionale, per quanto riguarda la sua ontologia in terza persona, esso consiste interamente di elementi neurobiologici; e la mancanza di condizioni causalmente sufficienti al livello psicologico ne è una conseguenza. Questo ci sembrerà meno complicato se riflettiamo sul fatto che il nostro bisogno di fermarci al livello dei neuroni è semplicemente una questione di pregiudizio. Se proseguiamo fino al livello quanto-meccanico, allora potrà sembrare meno sorprendente il fatto che vi sia un'assenza di condizioni causalmente sufficienti.

Sperry utilizza talvolta un esempio di causazione “dall’alto verso il basso” che un tempo pensavo fosse debole ma che ora mi sembra illuminante. Pensate ad una singola molecola in una ruota che sta girando. L’intera struttura della ruota e i suoi movimenti determinano i movimenti della molecola, anche se la ruota è costituita da tali molecole. E ciò che è vero per una molecola è vero per tutte le molecole. I movimenti di ogni molecola sono influenzati dal sistema sebbene il sistema consista interamente di molecole. Il modo corretto per pensare a questa situazione non è tanto una causazione “dall’alto verso il basso” quanto una causazione di sistema. Il sistema, in quanto sistema, ha degli effetti causali su ogni elemento, anche se il sistema è costituito da quegli stessi elementi. Ora, analogamente, secondo l’ipotesi 2, il sistema in quanto sistema cosciente, può avere degli effetti sugli elementi individuali, i neuroni e le sinapsi, anche se il sistema è costituito da essi. Ogni molecola in un liquido è influenzata dalla liquidità del sistema, anche se in esso non ci sono altri oggetti a parte le molecole. Allo stesso modo, nel cervello cosciente, ogni neurone nelle porzioni coscienti del sistema, può essere influenzato dalla coscienza del cervello, anche se non ci sono altri oggetti a parte i neuroni (con le cellule gliali, e tutto il resto).

Così, se l’ipotesi 2 è corretta, dobbiamo supporre che la coscienza del sistema abbia degli effetti sugli elementi del sistema, anche se il sistema è composto da elementi, nello stesso modo in cui la solidità della ruota ha degli effetti sulle molecole anche se la ruota è composta da molecole. Fino a qui sembra che tutto funzioni, ma l’analogia tra la causazione di sistema della ruota e la causazione di sistema del cervello cosciente crolla in questo punto: il comportamento della ruota è totalmente determinato mentre il comportamento del cervello cosciente, secondo l’ipotesi 2, non lo è. Com’è possibile? In che modo esattamente la neurobiologia funzionerebbe sulla base di una tale ipotesi? Io non conosco la risposta a questa domanda ma sono colpito dal fatto che molte delle spiegazioni fornite in neurobiologia non postulano condizioni causalmente sufficienti antecedenti. Così, per esempio, per citare un caso famoso, il “potenziale di prontezza” (*readiness potential*) discusso da Deecke, Scheid, Kornhuber e Libet⁴ non è causalmente sufficiente a determinare l’azione conseguente, come è posto in evidenza dalla discussione di Libet su come la coscienza possa interferire nell’operazione del potenziale di prontezza. Paradossalmente,

⁴ L. Deecke, P. Scheid, H. H. Kornhuber, “La distribuzione del ‘potenziale di prontezza’, la positività premotoria e il potenziale motorio della corteccia cerebrale umana precedono il movimento volontario del dito”, *Experimental Brain Research*, vol.7, 1969, pp. 158-168; B. Libet, “Do We Have Free Will?”, *Journal of Consciousness Studies*, 6, n. 8-9, 1999, pp. 47-57.

questi esperimenti sono talvolta presentati come argomenti contro il libero arbitrio. Questa conclusione non mi sembra sia implicata dai dati disponibili e farò ora una breve digressione per descrivere il problema.

Questo è ciò che succede. Il soggetto forma un'intenzione precedente cosciente di muovere il suo dito (o di far schioccare il polso) o di qualcosa che si fa tante volte. Si tratta di una decisione libera, cosciente. Sulla base di ciò, egli muove il suo dito coscientemente tante volte, e prima di ogni movimento del dito c'è un'attivazione del cervello, nella forma del potenziale di prontezza, registrabile sul cranio. In questi casi il potenziale di prontezza precede la consapevolezza cosciente dell'intenzione-in-azione di circa 350 millisecondi. In che modo si suppone che questo costituisca una minaccia per il libero arbitrio? Libet descrive il caso come una sorta di petizione di principio quando afferma che "l'inizio dell'atto volontario libero sembra iniziare nel cervello inconsciamente, ben prima che la persona sappia coscientemente di voler agire" (p. 51). Le espressioni "inizio" e "sappia di voler agire" possono essere fuorvianti. Questo è un altro modo per descrivere il caso: il soggetto, coscientemente, adotta una condotta per il movimento delle dita e, di conseguenza, sa quale tipo di azione vuole compiere quando prede quella decisione. Il cervello inconsciamente diviene pronto per ogni movimento prima dell'inizio cosciente del movimento. Nessuno, per quel mi è noto, sostiene che l'attivazione non sia relazionata alla decisione cosciente precedente; né qualcuno sostiene che l'attivazione sia una causa sufficiente per determinare il successivo movimento volontario del dito. La descrizione di Libet porta essa stessa all'interpretazione che il potenziale di prontezza segni l'inizio dell'azione. Ma questo non è vero. Trascorrono normalmente circa 350 millisecondi tra il potenziale di prontezza e l'inizio dell'intenzione-in-azione e altri 200 millisecondi prima dell'inizio del movimento corporeo. In ogni caso, per quanto sappiamo dai dati disponibili, l'avere luogo del potenziale di prontezza non è causalmente sufficiente per il compimento dell'azione. Da quanto mi risulta non conosciamo sufficientemente l'intera neurobiologia dell'azione intenzionale per avere una teoria completa del ruolo del potenziale di prontezza nella causazione dell'azione. Ma sembra chiaramente prematuro assumere che l'esistenza del potenziale di prontezza dimostri in ogni caso che noi non abbiamo il libero arbitrio.

Più interessanti sono i casi in cui il corpo inizia effettivamente a muoversi prima che il soggetto sia cosciente di una qualsiasi intenzione-in-azione di muoversi. Esempi famosi sono costituiti dal corridore che inizia a correre prima di aver sentito coscientemente lo

sparo della pistola, e dal giocatore di tennis il cui corpo inizia a muoversi verso la palla che sta arrivando prima di aver potuto coscientemente registrare il volo della palla nel suo sistema visivo. In entrambi i casi il corpo inizia effettivamente a muoversi prima che il soggetto sia consapevole, a livello cosciente, dello stimolo che dà il via al movimento. Nessuno di questi casi, tuttavia, mette in pericolo l'idea che in ogni caso noi abbiamo delle libere azioni volontarie. In entrambi i casi il soggetto, come risultato di un allenamento e di una pratica ripetuti, ha dei percorsi neurali ben stabiliti che sono attivati da stimoli percettivi che precedono l'inizio dell'attività della coscienza. Per porre questo punto in maniera esplicita: il soggetto sta giocando a tennis o correndo una gara per sua libera scelta, e se egli è piuttosto abile in queste attività il suo corpo deve essere capace di muoversi in certe situazioni chiave prima che egli sia coscientemente consapevole dello stimolo che provoca il suo movimento. La tentazione è di assimilare tutti questi casi – sia il potenziale di prontezza che l'atleta allenato – alla classe dei movimenti “riflessi” in cui l'agente effettivamente non ha alcun libero arbitrio. Così, per esempio, se accidentalmente tocco una stufa rovente io ritirerò la mia mano prima di sentire il dolore. Qui, mi sembra, le condizioni precedenti sono sufficienti a causare l'inizio dell'azione. Ma gli altri casi sono piuttosto differenti da questo. Sia nel caso del potenziale di prontezza che in quello dell'atleta allenato, i movimenti dipendono dal mio avere un'intenzione precedente cosciente – di muovere il mio dito, di giocare a tennis, di correre una corsa, ecc. – e io posso cancellare quell'intenzione in qualsiasi momento. Nel caso della stufa rovente non c'è nessuna intenzione precedente e io *non* avrei potuto *non* muovere la mia mano.

Proseguiamo nella nostra indagine. In che modo dobbiamo pensare alle relazioni tra i microelementi e la caratteristica di sistema della coscienza? Per le forme passive di coscienza, come la percezione, la totalità delle caratteristiche dei microelementi deve essere sufficiente, in ogni dato momento, a determinare lo stato cosciente di quel momento. Che dire della coscienza volitiva, il tipo di coscienza in cui esiste il divario? Mi sembra che in questo caso non valga lo stesso principio. La totalità delle caratteristiche per quanto riguarda i microlivelli rilevanti – neuroni, sinapsi, microtuboli o altro – sarebbe sufficiente a fissare unicamente lo stato cosciente a quel momento, inclusa la coscienza volitiva. Se dovessimo rinunciare a questo principio mi sembra che dovremmo accettare una qualche forma di dualismo. Dovremmo pensare ad una coscienza che si libera della sua base neurobiologica. Dovremmo rinunciare anche alle forme più ingenue di sopravvenienza, all'idea che ad ogni

cambiamento della coscienza debba corrispondere un cambiamento della neurobiologia.⁵ Il punto sul quale dobbiamo continuare ad insistere è che la coscienza non è una caratteristica ulteriore del cervello. Essa è semplicemente uno stato del sistema di neuroni così come la solidità della ruota non è un elemento ulteriore della ruota in aggiunta alle molecole. È semplicemente uno degli stati in cui si trovano le molecole.

Ma quando insistiamo sul fatto che le caratteristiche di sistema devono essere fissate unicamente dagli elementi del sistema noi non stiamo, per questo motivo, rinunciando al libero arbitrio, poiché il divario si estende nel tempo. Il divario non si trova tra lo stato dei miei neuroni in questo momento e il mio stato cosciente in questo momento; il divario si trova tra ciò che sta accadendo in questo momento nella parte volitiva cosciente dell'intero sistema cerebrale e ciò che accadrà successivamente.

Inoltre, si noti che ipotizzando una sequenza causale che non sia manifestata da condizioni causali sufficienti ad ogni livello noi non postuliamo la casualità. Perché no? Ricordiamo che ho sostenuto che dovremmo pensare alla coscienza come ad un campo cosciente unificato; e l'esperienza della volizione cosciente è un aspetto cruciale di quel campo cosciente. L'ipotesi che, fino a questo punto, viene suggerita dalla discussione è che noi dovremmo pensare all'agire razionale come ad una caratteristica di quel campo cosciente totale. Abbiamo osservato che, a livello psicologico, l'agire razionale può offrire delle spiegazioni causali dei fenomeni che non presentano una forma deterministica. Se l'agire razionale è realizzato in strutture neurobiologiche che hanno anche queste proprietà, che costituiscono esse stesse la struttura sottostante dell'agire razionale, allora ai processi neurobiologici mancherebbero le condizioni causalmente sufficienti, ma non per questo diverrebbero dei processi casuali. Essi sarebbero guidati dallo stesso agire razionale, che opera come una caratteristica del sistema.

L'ipotesi del divario è allora un'ipotesi neurobiologica che può essere ridotta a questo: il campo unificato della coscienza è un fenomeno biologico come qualsiasi altro. Esso è interamente spiegato dai processi neurobiologici. Tra questi processi ve ne sono alcuni che causano e realizzano la coscienza volitiva, la coscienza della deliberazione, della

⁵ Non sono un sostenitore del concetto di sopravvenienza. Il suo uso acritico è un segno di confusione filosofica, poiché il concetto oscilla tra la sopravvenienza causale e la sopravvenienza costitutiva. Ma noi vogliamo preservare l'idea ingenua che ogni cambiamento della coscienza debba essere contrassegnato da un cambiamento di tipo neurobiologico. Si veda il mio *Rediscovery of the Mind*, Cambridge, MA, MIT Press, 1992, pp. 124-126, per ulteriori discussioni, tr. it. *La riscoperta della mente*, Boringhieri, Torino, 1994, pp. 139-141.

scelta, della decisione e dell'azione. Date certe assunzioni riguardanti la natura di questi processi, la loro esistenza richiede un sé. Il sé non è un'entità del campo, ma determina un insieme di vincoli formali sulle operazioni del campo (come abbiamo visto nel capitolo 3). Espresso secondo i termini del nostro esempio, il fenomeno neurobiologico della libertà della volontà equivale a questi tre principi:

1. Ad ogni punto nel tempo t_1 lo stato cosciente totale del cervello, inclusa la coscienza volitiva, è interamente determinato dal comportamento dei microelementi rilevanti.
2. Lo stato del cervello a t_1 non è causalmente sufficiente a determinare lo stato del cervello a t_2 e t_3 .
3. Lo spostamento dallo stato t_1 allo stato a t_2 e t_3 può essere spiegato solamente attraverso le caratteristiche dell'intero sistema, specificamente dalle operazioni del sé cosciente.

Un modo per apprezzare la differenza tra queste due ipotesi è quello di applicare entrambe alla nostra finzione fantascientifica del robot immaginario, la Creatura, che avevamo costruito nel capitolo 5. In quel capitolo avevamo immaginato di aver costruito un robot cosciente, e che esso avesse l'esperienza del divario e le nostre esperienze. Ma ora chiediamoci in che modo potremmo pensare al libero arbitrio in quanto problema di ingegneria, come ad un problema di progettazione che collega la coscienza e la tecnologia. Se costruiamo il robot in accordo con l'ipotesi 1, costruiremo una macchina completamente deterministica; infatti, potremmo costruirla secondo i modelli tipici delle scienze cognitive dei sistemi computazionali, sia sistemi tradizionali che connessionisti. La macchina sarebbe progettata per ricevere degli *input* in forma di stimolazioni sensoriali, processerebbe questi ultimi secondo il suo programma e il suo database, e produrrebbe degli *output* in forma di movimenti muscolari. Per una macchina del genere, la coscienza potrebbe esistere, ma non giocherebbe alcun ruolo causale o esplicativo nel comportamento del sistema. Avendo costruito un sistema completamente deterministico, potremmo organizzarlo in modo tale che esso potrebbe avere, tramite la causazione inferiore-superiore, delle esperienze coscienti che corrispondono ai livelli delle sue operazioni di livello inferiore. Potrebbe soffrire di ansia e di indecisione al livello superiore, ma tutto questo sarebbe epifenomenico. Il meccanismo di livello inferiore determinerebbe completamente il comportamento conseguente del sistema.

Potremmo in effetti avere tutte queste caratteristiche e il sistema potrebbe tuttavia non essere prevedibile, poiché noi potremmo inserire qualche elemento casuale nel suo meccanismo che renderebbe il suo comportamento imprevedibile, anche se la coscienza fosse ancora epifenomenica. La coscienza esisterebbe ma sarebbe completamente inutile.

Secondo l'ipotesi 2 avremmo un tipo di problema ingegneristico completamente differente. Per l'ipotesi 2 l'intera organizzazione del campo cosciente unificato funziona essenzialmente all'interno delle capacità operazionali del sistema. La struttura e il comportamento dei microelementi, ad ogni dato momento, è sufficiente a determinare la caratteristica della coscienza in quell'istante, ma non è sufficiente a determinare lo stato successivo del sistema. Lo stato successivo del sistema è determinato solamente dal prendere decisioni coscienti, che è una caratteristica dell'intero sistema. In quanto problema di ingegneria, non ho idea di come potremmo proseguire nella costruzione, ma allora, allo stato attuale, non abbiamo alcuna idea di come sia possibile costruire un robot cosciente.

Una volta riconosciuta la realtà psicologica del divario, mi sembra che queste siano le due forme di spiegazione del comportamento umano più attendibili. Per prima cosa, l'indeterminismo psicologico coesiste con il determinismo neurobiologico. Se questa tesi è vera la vita razionale libera è interamente un'illusione. L'altra possibilità è che all'indeterminismo psicologico corrisponda un indeterminismo neurobiologico. Ho cercato di dimostrare che questa è, perlomeno, una possibilità empirica. Non ho idea se una di queste due ipotesi si rivelerà essere vera. Forse si rivelerà vera una terza possibilità che non possiamo nemmeno immaginare. Queste sono le due ipotesi a cui possiamo giungere se seguiamo in modo inflessibile la linea di indagine suggerita sia da quanto sappiamo dalla nostra esperienza, sia da ciò che sappiamo riguardo al nostro cervello.

Francamente, non trovo nessuna delle due ipotesi intellettualmente affascinante. L'ipotesi 1 è confortante in quanto ci permette di trattare il cervello nello stesso modo in cui trattiamo qualsiasi altro organo. Noi consideriamo il cervello come un sistema completamente deterministico, come il fegato o il cuore. Ma l'ipotesi 1 non è consolante rispetto a ciò che sappiamo riguardo all'evoluzione. Secondo questa ipotesi, c'è un sistema cosciente enormemente elaborato e dispendioso, il sistema del prendere decisioni razionali, che non gioca nessun ruolo causale nel comportamento dell'organismo, poiché il comportamento è interamente fissato al livello inferiore. Secondo questa prospettiva, non ci sarebbe nessun vantaggio selettivo nell'avere un sistema decisionale cosciente, razionale,

che è il risultato di un lungo periodo di evoluzione, che è estremamente dispendioso, dal punto di vista biologico, e che occupa uno spazio enorme nelle nostre esperienze coscienti. Inoltre, l'illusione del prendere decisioni razionali, sulla base di questa ipotesi, non sarebbe come le altre illusioni, che hanno in effetti un vantaggio selettivo. Così, per esempio, assumendo che il colore sia un'illusione sistematica, c'è non di meno un enorme vantaggio selettivo in un organismo che ha la capacità di discriminare gli oggetti sulla base del colore. Ma secondo l'ipotesi numero 1 il prendere decisioni razionali coscienti non porterebbe nessun tipo di vantaggio selettivo.

Ma l'ipotesi numero 2, del resto, non è confortante per la nostra attuale concezione della biologia. Il problema non è che l'ipotesi 2 ci chiede di pensare alla coscienza in relazione al suo ruolo causale “dall'alto verso il basso” nel comportamento dei microelementi, poiché, essendo una caratteristica di sistema, la coscienza funziona come qualsiasi altra caratteristica di sistema. Alla fine, quando parliamo del fatto che la coscienza influisce sugli altri elementi, noi stiamo parlando effettivamente solo di come gli elementi influiscono gli uni sugli altri poiché la coscienza è interamente una funzione del comportamento degli elementi. Allo stesso modo, quando parliamo del fatto che il comportamento della ruota influisce sulle molecole, stiamo semplicemente parlando di come le molecole si influenzano reciprocamente. Così il problema dell'ipotesi 2 non è il fatto che essa implica una causazione “dall'alto verso il basso” della coscienza. Questo è un problema semplice. Il problema complesso è vedere in che modo la coscienza del sistema possa fornire al sistema stesso un'efficacia causale che non sia deterministica. E non è un aiuto sufficiente sostenere che potremmo accettare la casualità delle spiegazioni non deterministiche della meccanica quantistica. Si suppone che la razionalità cosciente non erediti la casualità della meccanica quantistica. Piuttosto, si suppone che la razionalità cosciente sia un meccanismo causale che agisce causalmente, sebbene non sulla base di precedenti condizioni causali sufficienti. Secondo alcune spiegazioni, in effetti, una delle funzioni della cellula è di superare l'instabilità dell'indeterminatezza quantistica ai livelli subcellulari.

Non ho cercato di risolvere il problema della libertà della volontà arbitrio, ma ho semplicemente provato ad esporre in che cosa consista esattamente il problema, e ad indicare quali potrebbero essere le linee più attendibili per una sua possibile soluzione.

POSTFAZIONE

LA TEORIA NON CAUSALE DELLA RAZIONALITÀ DI JOHN R. SEARLE

*Eddy Carli**

Il tema della razionalità pratica ha un'origine antica, che risale ad Aristotele: egli per primo teorizzò la necessità di pensare ad una forma peculiare di razionalità, caratterizzante l'agire pratico dell'uomo e distinta dalla razionalità teoretica. Aristotele sviluppò una teoria della razionalità pratica a partire dalla distinzione tra sapere teoretico e sapere pratico, distinzione che era innanzitutto rivolta contro l'unificazione di virtù e sapere pratico operata da Socrate e contro l'indistinzione tra i due diversi saperi presente in Platone¹.

Per Aristotele la "filosofia prima", o "sapienza" (*sophia*), costituiva il sapere supremo, ma, essendo una scienza puramente teoretica, come la fisica e la matematica, non era adeguata a descrivere e spiegare i fenomeni che comprendono le azioni degli uomini; questo doveva essere compito di un altro tipo di sapere, il sapere "pratico", che ha per oggetto non gli enti fisici o i concetti filosofici astratti, ma appunto le azioni. Le scienze teoretiche hanno per oggetto realtà che non dipendono dagli uomini, poiché esse studiano la forma delle sostanze mobili che hanno il principio del loro movimento in se stesse (le sostanze naturali), e tale forma può essere considerata sia in quanto unita alla materia, nel caso della fisica, sia separata dalla materia, nel caso della matematica; oppure, esse possono avere come oggetto di studio le sostanze immobili, e questo è il caso della "filosofia prima", detta perciò anche "scienza teologica" o, come poi sarà chiamata, "metafisica". Se le scienze teoretiche culminano nella "sapienza" (*sophia*), il sapere pratico culmina invece nella "saggezza" o "prudenza" (*phronesis*), e ciò che differenzia i due ambiti, teoretico e pratico, sono i diversi tipi di dimostrazione che di essi è possibile fornire. Il metodo della filosofia pratica si distingue da quello delle scienze teoretiche, in cui le dimostrazioni valgono "sempre": nell'ambito della razionalità pratica le premesse sono valide "per lo più" e così le conclusioni². Questo non impedisce tuttavia, secondo Aristotele, che anche il sapere pratico (come l'etica e la politica) sia una scienza, cioè un "abito dimostrativo", una capacità di fare dimostrazioni, perché per Aristotele sono scienze sia quelle le cui dimostrazioni valgono sempre, cioè approdano a conclusioni necessarie, sia quelle le cui dimostrazioni valgono per lo più, cioè approdano a conclusioni valide nella maggior parte dei casi³.

* Ringrazio Enrico Berti per le preziose indicazioni relative alla filosofia pratica di Aristotele e John R. Searle per le discussioni sulla razionalità pratica.

¹ Si veda E. Berti, "Le forme del sapere nel passaggio dal premoderno al moderno" in Id. (a cura) *La razionalità pratica. Modelli e problemi*, Marietti, Genova 1989, pp. 15-41.

² Aristotele, *Etica Nicomachea*, I 1094b 20-25, Laterza, Roma-Bari 1973 (nuova edizione, 1999).

³ "Ogni sillogismo si sviluppa o attraverso premesse necessarie o attraverso premesse esprimenti qualcosa che avviene per lo più. Inoltre, se le premesse sono necessarie, anche la conclusione risulta necessaria; se invece le premesse esprimono ciò che avviene per lo più, anche la conclusione esprimerà qualcosa di simile". Aristotele, *Analitici posteriori*, I (A) 30, 87b, 20-25, Laterza, Roma-Bari 1973, 1994.

Non tutte le scienze hanno infatti carattere di assolutezza, e Aristotele ne è ben consapevole quando afferma che ogni scienza ha per oggetto o ciò che è sempre o ciò che è per lo più e che la maggior parte delle scienze ha per oggetto ciò che è per lo più⁴ e che è dunque possibile avere scienza di ciò che è per lo più⁵.

Per la razionalità pratica egli propone un particolare modello logico, il sillogismo pratico, il quale è appunto un tipo di ragionamento diverso da quello scientifico poiché non conduce *necessariamente* a conclusioni certe e indubitabili. Nel sillogismo pratico c'è una premessa universale, costituita dall'indicazione del fine, una premessa particolare, costituita dall'indicazione dei mezzi e una conclusione, costituita dall'azione stessa⁶. Diversamente dal sillogismo scientifico, nel sillogismo pratico la conclusione non segue necessariamente dalle premesse, e proprio questa caratteristica è ciò che rende peculiare tale sillogismo e consente, secondo Aristotele, di eleggerlo a modello logico dell'azione umana.

Con Aristotele abbiamo dunque la prima grande differenziazione tra sapere teoretico e sapere pratico e l'elaborazione di un modello esplicativo dell'azione umana che, se pure con alcune limitazioni di carattere logico ed epistemologico,⁷ rimane ancora oggi una valida analisi logica dell'azione umana. Esso costituisce, almeno in parte, quello che John Searle chiama il "Modello Classico della razionalità": modello che egli considera il primo e più radicale bersaglio critico della sua intera argomentazione sulla razionalità pratica, o "razionalità nell'azione" (*rationality in action*), e che tuttavia, nella sua formulazione originaria, quella aristotelica appunto, presenta notevoli affinità con la teoria della razionalità pratica che Searle sviluppa in questo suo nuovo libro, *Razionalità nell'azione*, interamente dedicato a tale analisi.

Searle, come Aristotele, riconosce la necessità di un'indagine della "razionalità nell'azione" distinta da quella della razionalità teorica, o teoretica, e sostiene che, pur non essendoci una logica deduttiva del ragionamento pratico analoga a quella del ragionamento scientifico, vi è tuttavia una possibilità di descrivere l'azione umana e la razionalità pratica che la determina attraverso una logica del "per lo più" che richiama il sillogismo pratico aristotelico.

L'analisi di Searle si sviluppa poi in modo originale, collegandosi agli scritti precedenti del filosofo americano: nella teoria della razionalità di Searle si ritrovano la filosofia del linguaggio e la filosofia della mente che hanno segnato il suo percorso intellettuale. La teoria degli atti linguistici (*Speech acts*) e dell'intenzionalità, l'irriducibilità della coscienza e il naturalismo biologico si rivelano gli elementi essenziali che consentono a Searle di elaborare una originale teoria "linguistica" della razionalità in azione, di eredità wittgensteiniana, a partire da una critica radicale del modello tradizionale della razionalità pratica secondo cui essa non implicherebbe alcun tipo di conoscenza, almeno non nel senso in cui si giunge alla conoscenza nell'ambito della ragione teoretica. E qui il

⁴ Aristotele, *Metaph.* K 8, 1065a 4-6.

⁵ Si veda M. Mignucci, *La teoria aristotelica della scienza*, Sansoni, Firenze, 1965, cap. III.

⁶ Uno degli esempi di sillogismo pratico che Aristotele riporta nell'*Etica Nicomachea* è il seguente: le carni leggere producono la salute (premessa universale), le carni di uccello sono leggere (premessa particolare), dunque mangiamo carni di uccello (conclusione). *Eth. Nic.* VI 8, 1141b 18-21.

⁷ Nell'ambito della filosofia analitica, la rilevanza, e i limiti, del sillogismo pratico aristotelico come modello descrittivo dell'azione intenzionale sono stati sottolineati soprattutto da Elizabeth M. Anscombe, *Intention*, Basil Blackwell, Oxford 1957, da Georg. H. von Wright, *Explanation and Understanding*, Cornell University Press, Ithaca, New York 1971, tr. it. *Spiegazione e comprensione*, Il Mulino, Bologna 1977 e da Donald Davidson, *Actions Reasons and Causes*, 1963, tr. it. "Azioni, ragioni e cause", in *Azioni ed eventi*, il Mulino, Bologna 1992.

bersaglio critico di Searle diviene esplicitamente Immanuel Kant⁸ e il suo rifiuto a riconoscere alla filosofia pratica la capacità di sviluppare una vera forma di conoscenza, se pure diversa da quella della filosofia teoretica: la "ragione pratica" di Kant, diversamente dalla "razionalità pratica" di Aristotele, di propriamente razionale ha l'universalità e la necessità, cioè il carattere di *a priori*, ma non implica nessuna forma di conoscenza.

Va osservato che l'attacco di Searle a Kant è alquanto diverso da quelli usuali; infatti, la critica non considera il fatto che nella filosofia pratica di Kant il modello della spiegazione teleologica ha una rilevanza essenziale⁹, né il fatto che per Kant il problema della "ragion pratica" si pone sul piano della morale ed investe principalmente il tema della volontà. "Pratico" (*praktisch*) è per Kant ciò che attiene alla determinazione della volontà e, poiché le condizioni della libertà sono empiriche, la ragione, in ambito pratico, può avere solo un uso regolativo, ovvero di tipo etico-pratico. Secondo Kant, dal punto di vista della morale, non sono tanto le azioni ad essere rilevanti (in quanto realizzazioni effettive di propositi e intenzioni): ciò che conta sono esclusivamente le intenzioni (*Gesinnungen*), sia che si realizzino sia che non si realizzino in azioni¹⁰. Searle sostiene invece che *solamente* nell'ambito della razionalità pratica, intesa come razionalità che conduce all'azione, è possibile giungere a conoscere e a comprendere i comportamenti intenzionali, coscienti che caratterizzano le vite degli esseri umani, e tale conoscenza non può che avvenire a partire dall'analisi del linguaggio, dei suoi usi e della sua grammatica.

La critica al "modello classico" della razionalità

Per Searle, il problema della razionalità è un problema filosofico persistente che ha origine antica, come abbiamo già osservato, e che nella modernità, in particolare con Hume e con Kant, si caratterizzerà essenzialmente come razionalità teorica, contrapposta alla razionalità pratica. Ma, sostiene Searle, se accettassimo tale "Modello Classico" della razionalità saremmo costretti a pensare che la razionalità umana corrisponda alla capacità di simbolizzazione e che essa sia allora semplicemente una versione più complessa della razionalità di alcune scimmie particolarmente evolute e opportunamente addestrate, le quali, pur presentando una certa capacità di simbolizzazione, non hanno tuttavia adeguate capacità linguistiche e si rivelano così incapaci di agire razionalmente, nel senso degli esseri umani. Come dimostrano esperimenti su scimpanzé famosi come Washoe e Lana¹¹, la semplice capacità di simbolizzazione non è di per sé sufficiente a giustificare l'intera gamma dei processi del pensiero razionale. Gli sforzi per insegnare agli scimpanzé ad usare simboli linguistici – afferma Searle¹² – hanno portato a dei risultati piuttosto ambigui. Ciò che è necessario è la capacità di disporre di certi tipi di rappresentazioni linguistiche e tali rappresentazioni mancano agli animali non umani, anche a quelli più evoluti, come gli scimpanzé.

⁸ Il testo di Kant cui Searle fa riferimento è la *Fondazione della metafisica dei costumi*, tr. it. in *Scritti morali*, Utet, Torino 1970.

⁹ Kant, I. *Critica della ragion pratica*, tr. it. Laterza, Roma-Bari 1997.

¹⁰ Si veda l'introduzione di S. Landucci alla *Critica della ragion pratica*, tr. it. Laterza, Roma-Bari 1997, p. XIII.

¹¹ Esperimenti riportati dallo stesso Searle nel cap. I di questo volume e che si riferiscono al testo di Wolfgang Köhler, *L'intelligenza nelle scimmie antropoidi*, presentazione e traduzione di Guido Petter, Giunti-Barbera, Firenze 1968.

¹² *Ibid.*, cap. I.

A partire dalla pragmatica del linguaggio e dalla teoria degli atti linguistici, Searle elabora una teoria della razionalità in cui svolgono un ruolo essenziale le ragioni, intese come motivazioni che conducono all'azione, in contrapposizione alle cause, e inoltre le intenzioni e il divario (*gap*), ovvero quella sorta di scarto tra l'intenzione e l'azione che è ciò che maggiormente caratterizza l'agire umano, quello che nella tradizione filosofica è chiamato "libero arbitrio". Secondo Searle, il "Modello Classico" della razionalità non considera tali concetti come caratteri essenziali della ragione umana e proprio per questo incorre in quegli errori che egli, nel primo capitolo del libro, elenca come i 6 errori principali del "modello classico". Questi sono:

1. Le azioni, se razionali, sono *causate* da credenze e desideri.
2. La razionalità è una questione di osservazione delle regole, le regole speciali che costituiscono la distinzione tra pensiero e comportamento razionale e irrazionale.
3. La razionalità è una facoltà cognitiva separata.
4. Casi apparenti di debolezza della volontà, di *akrasia*, possono sorgere solo nei casi in cui c'è qualcosa di sbagliato negli antecedenti psicologici dell'azione.
5. La ragione pratica deve iniziare con un inventario degli scopi primari dell'agente, incluse le mete e i desideri fondamentali, gli obiettivi e i propositi; e questi non sono essi stessi soggetti a costrizioni razionali.
6. L'intero sistema della razionalità funziona solo se l'insieme dei desideri primari è consistente.

Per Searle ognuna di queste affermazioni è falsa, ed egli dedica gran parte del libro a dimostrare perché le tesi del Modello Classico sono errate per giungere poi a sviluppare un'originale teoria della razionalità, basata sull'analisi del linguaggio, nella quale svolgono un ruolo primario le ragioni e i desideri che determinano le azioni volontarie e intenzionali.

Il Modello Classico può descrivere dei casi particolari ma non fornire una teoria generale del ruolo della razionalità nel pensiero e nell'azione: innanzitutto perché nel caso normale di un'azione razionale l'insieme di credenze e desideri antecedente non è causalmente sufficiente a determinare l'azione. Questo è per Searle un presupposto del processo di deliberazione ed è indispensabile per l'applicazione della razionalità. Ciò di cui il Modello Classico non tiene conto è il divario (o "libero arbitrio") tra le "cause" dell'azione, le credenze e i desideri, e l'"effetto", ovvero l'azione. Tale divario è secondo Searle ciò che impedisce di fornire un modello della razionalità pratica nei termini di una spiegazione causale di carattere nomologico-deduttivo, secondo il modello di spiegazione scientifica fornito da Hempel, per il quale è possibile ricondurre fenomeni ed eventi ad un causalismo che si esprime in leggi rigorose¹³. Searle mette radicalmente in discussione la possibilità di un tale modello, poiché esso impedirebbe l'esercizio del libero arbitrio che sta a fondamento della razionalità pratica: se le *ragioni* fossero *cause* delle azioni e se esse fossero descrivibili nei termini di leggi rigorose in senso stretto, verrebbe meno la capacità stessa di agire liberamente.

¹³ In *Philosophy of Natural Science*, Prentice-Hall, N.J., 1966, Hempel dà la seguente definizione di spiegazione nomologico-deduttiva: "Le spiegazioni nomologico-deduttive soddisfano la richiesta di rilevanza esplicativa nel senso più forte possibile: l'informazione di carattere esplicativo che esse forniscono implica la deduzione dell'*explanandum* ed offre così delle ragioni logicamente conclusive del perché ci si attenda il fenomeno dell'*explanandum*", p. 52. Va tenuto conto tuttavia che lo stesso Hempel sosteneva che il modello di legge causale può essere anche di tipo probabilistico.

Per impegnarci in una decisione razionale dobbiamo presupporre l'esistenza del libero arbitrio poiché esso è presente in qualsiasi attività razionale. Non possiamo evitare tale presupposizione, poiché persino il rifiuto ad impegnarsi in una decisione razionale ci risulta intelligibile come rifiuto soltanto se lo consideriamo come un esercizio della libertà.

Cause e ragioni dell'azione

Searle propone una teoria non-causale della razionalità, secondo la quale non solo non è possibile descrivere i comportamenti umani nei termini di cause e di leggi rigorose, ma che afferma che *le ragioni non sono cause delle azioni intenzionali*. Già Donald Davidson¹⁴ aveva posto il problema della causazione psicofisica negando l'esistenza di leggi rigorose capaci di descrivere i fenomeni intenzionali, ma egli sosteneva che *le ragioni sono le cause dell'agire intenzionale*, poiché il suo obiettivo era quello di conciliare la causalità di carattere nomologico-esplicativo alla Hempel, con l'anticausalismo che nega la possibilità di spiegare l'azione tramite il richiamo a leggi rigorose. Per Davidson dunque, sebbene le azioni siano *causalmente* provocate da credenze, desideri e intenzioni esse non possono essere ricondotte a leggi rigorose (*strict laws*).

Searle radicalizza la tesi di Davidson poiché afferma non solo che le leggi psicofisiche non sono possibili ma che le cause non sono equivalenti alle ragioni¹⁵: egli eredita in questo senso le tesi di Ludwig Wittgenstein¹⁶ e della sua allieva prediletta Elizabeth Anscombe¹⁷ secondo cui, a partire dall'analisi del linguaggio ordinario, si giunge alla necessità di porre una differenziazione tra le *ragioni* e le *cause* di un'azione intenzionale. Per Wittgenstein, dobbiamo sostituire le "cause" con le "ragioni" e introdurre in questo modo una concezione "grammaticale" di tali concetti. "Dare una ragione di qualcosa che si è fatto o si è detto significa mostrare una via che porta a questa azione. In alcuni casi, questo significa descrivere la via che si è percorsa; in altri invece, significa *descrivere una via che porta a questa azione* e che è conforme a certe regole accettate"¹⁸. Anscombe, svilupperà in modo più sistematico le indicazioni di Wittgenstein fino a sostenere una esplicita teoria non-causale dell'azione intenzionale, nella quale un ruolo fondamentale è svolto dalla ragione pratica e, in particolare, dalla teoria del sillogismo pratico aristotelico, che secondo l'autrice risponde in modo adeguato all'esigenza di una spiegazione logica dell'azione intenzionale¹⁹.

Searle, sebbene senza citare esplicitamente Anscombe, ne utilizza il metodo di indagine e i criteri per descrivere l'azione intenzionale, ovvero: la "domanda perché?" (*Question Why*)²⁰, in senso "speciale" cioè come domanda cui è possibile dare una risposta solamente fornendo la *ragione* che ha

¹⁴ D. Davidson, *Actions, Reasons and Causes*, in "The Journal of Philosophy", 60, 1963, tr. it. "Azioni, ragioni e cause", in *Azioni ed eventi*, Il Mulino, Bologna 1992, 2000.

¹⁵ "Si può ottenere una ragione enunciando una causa, ma questo non implica che la ragione e la causa siano la stessa cosa", J. Searle, presente volume, cap. 4, p. 107

¹⁶ L. Wittgenstein, *Philosophical Investigations*, Blackwell, Oxford 1958, (ed. or. *Philosophische Untersuchungen*, 1953) tr. it. *Ricerche filosofiche*, Einaudi, Torino 1967; *The Blue and Brown Book*, 1958, tr. it. *Libro blu e libro marrone*, Einaudi, Torino 1983.

¹⁷ Anscombe, G.E.M. *Intention*, Basil Blackwell, Oxford 1957.

¹⁸ *Blue Book*, tr. it. p. 23.

¹⁹ Anscombe definisce il sillogismo pratico "una delle migliori scoperte di Aristotele" ("*one's of Aristotle's best discoveries*"), *Intention*, § 33, pp. 57-58.

²⁰ Anscombe, *Intention*, §§ 5-20; Searle, capitoli 3 e 4 del presente volume.

provocato una certa azione intenzionale; l'individuazione dell'intenzione con cui (*with which*) e tramite cui (*by which*)²¹ un'azione viene compiuta; e la descrizione delle condizioni sotto le quali (*under which*)²² una particolare azione ha luogo. E, in particolare, ne utilizza l'argomento a favore della connessione logica tra ragioni e azione (*the logical connection argument*)²³: tra ragioni e azioni non c'è nessuna connessione causale perché una relazione causale può essere scoperta solo induttivamente, mai attraverso un'analisi concettuale, mentre la relazione tra ragione e azione è di carattere logico-concettuale. Poiché c'è una relazione logica tra l'intenzione *I* di un agente *X* di fare *A* e il suo fare *A*, *A* non può essere un effetto di *I*.

Se una persona corre verso un treno in partenza, possiamo attribuirgli l'intenzione di salire su quel treno, ma quell'intenzione, che poniamo all'origine della sua azione di correre verso il treno, non è indipendente dalla nostra descrizione dell'azione. L'azione viene descritta come azione intenzionale solamente se è possibile descrivere le *ragioni per le quali* essa viene compiuta: si corre intenzionalmente verso il treno se si vuole prenderlo, se si è formulata l'intenzione di voler raggiungere Venezia entro le ore 9, salendo su quel treno, se si ha la credenza che quell'oggetto sia un treno, ecc. Tra il contenuto proposizionale degli stati intenzionali che spiegano l'azione e la descrizione mediante la quale quell'azione viene descritta c'è in questo caso un nesso concettuale e non causale.

Le ragioni che determinano l'azione sono per Searle delle "entità che hanno una struttura fattiva"²⁴ e la spiegazione di un enunciato di ragione è un atto linguistico che consiste nel fornire delle ragioni. Un enunciato di ragione sarà una spiegazione solo se la ragione stessa si trova in una, o più d'una, delle relazioni esplicative delle cose di cui è una ragione. E sebbene a volte l'enunciato di ragione specifichi una causa, non ne segue che in tali casi la causa sia identica alla ragione, perché le ragioni sono sempre delle entità fattive e le cause sono tipicamente degli eventi, non dei fatti. I nessi che è possibile rinvenire sia negli atti linguistici che nelle relazioni tra intenzione e azione sono dei nessi logico-concettuali, non causali.

Lo sfondo della teoria della razionalità pratica di Searle è costituito dalla sua teoria degli atti linguistici (*speech acts' theory*)²⁵ e dalla sua teoria dell'intenzionalità²⁶: soltanto a partire da un'indagine pragmatica del linguaggio possiamo giungere a "pensare nei termini di nuovi giochi linguistici"²⁷ e solo da una teoria generale dell'intenzionalità possiamo sviluppare una teoria generale della razionalità. Filosofia del linguaggio e filosofia della mente si intersecano dunque e all'analisi della razionalità vengono applicati i criteri degli atti linguistici: desideri, intenzioni, credenze, volizioni sono gli atti mentali che caratterizzano la ragione pratica ed essi hanno "direzioni di

²¹ Anscombe, 1957, §§ 22-25.

²² Anscombe, 1957, §§ 22-30.

²³ Argomento di eredità humiana: per Hume, un evento non può essere causa di un altro se tra di loro c'è un nesso di carattere logico-concettuale, che è invece quanto accade nel caso delle ragioni e delle azioni. Causa ed effetto devono essere concettualmente irrelati.

²⁴ Si veda il presente volume, cap. 4, p. 107.

²⁵ J. Searle, *Speech Acts. An Essays in the Philosophy of Language*, Cambridge University Press, London 1969, tr. it. *Atti linguistici. Saggio di filosofia del linguaggio*, Boringhieri, Torino, 1976.

²⁶ J. Searle, *Intentionality. An Essay in the Philosophy of Mind*, Cambridge University Press, London 1983, tr. it. *Della intenzionalità. Un saggio di filosofia della conoscenza*, Bompiani, Milano 1985.

²⁷ Searle si riferisce in modo esplicito alla teoria dei "giochi linguistici" di Wittgenstein. In questo volume, cap. 3, p. 85-86.

adattamento" e "condizioni di soddisfazione" mente-a-mondo o mondo-a-mente così come le hanno gli atti linguistici e le azioni intenzionali.

Sebbene la teoria della razionalità di Searle, nel sostenere una prospettiva non-causale in cui le ragioni svolgono un ruolo essenziale in quanto nozioni distinte dalle cause, prosegua, come abbiamo cercato di dimostrare, la tradizione classica della filosofia analitica del linguaggio, in particolare con Wittgenstein e Anscombe, se ne distanzia poi in modo significativo elaborando una originale teoria della razionalità in cui si ritrovano le analisi sugli atti linguistici, l'intenzionalità, la mente e la coscienza.

L'impossibilità di una logica formale della razionalità pratica

Per Searle non è possibile giungere ad un modello logico-deduttivo dell'azione intenzionale e della razionalità pratica²⁸, quel modello che Anscombe individuava nel sillogismo pratico aristotelico, poiché esso risulta inadeguato a dare ragione di scelte conflittuali o di volontà debole, come quelle costituite dai casi di *akrasia*²⁹, in cui il soggetto pur formando un'intenzione fortissima di fare qualcosa alla fine non la fa. Searle critica Hare³⁰ e Davidson³¹ per i quali la volontà debole è una forma di irrazionalità e sostiene invece che è necessario pensare alla razionalità in senso più ampio in cui comprendere anche i casi di debolezza della volontà. L'argomentazione a favore dell'*akrasia* diviene per Searle un punto di forza per la critica al modello causale dell'azione razionale: il modello generale, cui fanno riferimento Hare e Davidson, è supporre che se gli antecedenti dell'azione sono strutturati razionalmente in un certo modo, allora l'azione seguirà per necessità causale. Ma questo contrasta con la tesi di Searle secondo cui le ragioni, le intenzioni, i desideri, la volontà non sono degli antecedenti causali dell'azione e da essi non segue dunque *necessariamente* l'azione. Proprio il fatto che da un'intenzione fortissima possa seguire un'azione contraria a tale intenzione dimostra, secondo Searle, che le azioni non accadono necessariamente, ma per lo più e che la debolezza della volontà è un elemento che caratterizza l'agire razionale.

Anna ha la ferma intenzione di perdere peso entro il prossimo mese, poiché desidera indossare l'abito rosso che aveva acquistato lo scorso anno al matrimonio della sorella, ma non resiste alla tentazione di mangiare del cioccolato e dunque, razionalmente, mangia del cioccolato. Anna compie dunque un'azione contraria alla sua intenzione e al suo desiderio primario, soddisfacendo tuttavia un desiderio secondario e compiendo un'altra azione, anch'essa tuttavia razionale. Dalla sua intenzione primaria non segue per necessità causale l'azione.

La tesi a favore dell'*akrasia* dimostra per Searle l'impossibilità di ottenere una logica formale della ragione pratica: nell'inferenza pratica l'agente che accetta le premesse non è impegnato a desiderare o intendere la conclusione, nello stesso modo in cui l'agente che accetta le premesse di un'inferenza teorica è impegnato a credere nelle conclusioni. La ragione pratica implica essenzialmente il giudizio

²⁸ Si veda il cap. 8 del presente volume.

²⁹ Si veda il cap. 7 del presente volume.

³⁰ R.M. Hare, *The Language of Morals*, Oxford, Oxford University Press, 1952, tr. it. *Il linguaggio della morale*, Ubaldini, Roma 1968

³¹ "How Is Weakness of the Will Possible?" in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, tr. it. "La debolezza della volontà", in *Azioni ed eventi*, il Mulino, Bologna 1992.

di desideri, intenzioni, ragioni e motivi in conflitto mentre la ragione teorica non implica essenzialmente il giudizio di credenze in conflitto.

La ricerca di una struttura logica deduttiva formale della ragione pratica è per Searle fuorviante. I modelli che sono stati elaborati hanno pochissime applicazioni e, se vengono predisposti per essere applicati nella vita reale, questo implica una banalizzazione della caratteristica essenziale della deliberazione pratica: la riconciliazione tra desideri conflittuali e ragioni per l'azione conflittuali e la formazione di desideri razionali sulla base di tale riconciliazione. Potremmo sempre costruire un modello deduttivo per qualsiasi parte del ragionamento; ma quando una caratteristica essenziale del ragionamento contiene sia p che non p – come quando l'agente vuole che p e vuole che non p , nel caso di Anna, ella vuole dimagrire e non vuole dimagrire (mangiando il cioccolato) – la logica deduttiva non è in grado di risolvere tali incoerenze.

La teoria della razionalità pratica di Searle, si dimostra dunque soltanto in parte uno sviluppo delle teorie di Wittgenstein e di Anscombe, delle quali accetta sostanzialmente il metodo di indagine, la grammatica e gli usi dei diversi giochi linguistici, ma non i contenuti epistemologici ed ontologici. Per Searle non è possibile una logica formale della ragione pratica e non è possibile disgiungere l'analisi filosofica, grammaticale e logica, da un'analisi scientifica. La comprensione e la spiegazione della razionalità pratica sono strettamente legate allo studio della struttura del cervello: è dai progressi delle neuroscienze che è possibile, secondo Searle, giungere ad ulteriori conoscenze nell'ambito della mente e della razionalità umana.

Razionalità, coscienza e cervello

La razionalità, come la coscienza e l'intenzionalità, sono per Searle fenomeni naturali così come lo sono la fotosintesi e la mitosi, i tramonti sul mare, le maree e le montagne; essi hanno una realtà oggettiva e sono dunque indagabili empiricamente. Nella sua forma filosofica il problema mente-corpo ha secondo Searle una soluzione piuttosto semplice, che si basa sulla neurobiologia del cervello e alla quale egli stesso ha dato il nome di "naturalismo biologico" – una sorta di naturalismo non-riduzionista che cerca di tenere assieme gli aspetti soggettivi e oggettivi della coscienza e del cervello umani: tutti i nostri stati coscienti sono causati da processi neuronali di livello inferiore del cervello, e sono essi stessi delle caratteristiche del cervello.

Le esperienze soggettive, qualitative del dolore nel campo cosciente totale sono causate da processi neurobiologici del cervello e dal resto del sistema nervoso centrale e sono essi stessi, in quanto elementi del campo unificato della coscienza, delle caratteristiche del sistema di neuroni e di altre cellule che costituiscono il cervello umano. Il campo unificato della coscienza è un fenomeno biologico come qualsiasi altro. Esso è interamente spiegato dai processi neurobiologici. Tra questi processi ve ne sono alcuni che causano e realizzano la razionalità pratica, costituita dalla coscienza volitiva, dalla coscienza della deliberazione, della scelta, della decisione e dell'azione.

Searle afferma che non è sufficiente riconoscere la realtà psicologica della razionalità pratica e del divario (*gap*) che la caratterizza, ma è necessario indagarne la realtà neurobiologica. Se la libertà umana esiste realmente, essa deve essere una caratteristica delle funzioni del cervello: la spiegazione

della coscienza volontaria si colloca dunque all'interno della spiegazione generale della coscienza e tale spiegazione è a sua volta inserita all'interno di una spiegazione delle funzioni cerebrali. Il "naturalismo biologico" comprende dunque anche la spiegazione della realtà del divario o della libertà umana: l'ipotesi del divario è un'ipotesi neurobiologica che può essere ridotta al campo unificato della coscienza, ed essa è un fenomeno naturale.

Una volta riconosciuta la realtà psicologica e neurobiologica della ragione pratica e del divario, è possibile sostenere la tesi secondo la quale all'indeterminismo psicologico, che caratterizza i fenomeni mentali, corrisponderebbe un indeterminismo neurobiologico³². Tesi che è per Searle perlomeno una possibilità empirica e che è in parte supportata dalle indagini svolte in campo neurobiologico da ricerche sperimentali come quelle di Gerald Edelman e dalla sua teoria scientifica della coscienza, fondata sulla teoria darwiniana della selezione dei gruppi neuronici (TSGN)³³. L'indeterminismo presente nella teoria neurobiologica della coscienza di Edelman è dovuto alla selezione di tipo darwiniano operata dai gruppi neuronici, le cui relazioni risultano pertanto di tipo causale, indeterminate.

Estendendo la tesi del "naturalismo biologico" alla razionalità pratica e al divario Searle afferma dunque la possibilità di una descrizione scientifica, se pure non deterministica, della mente e della razionalità e conferma, come già ha fatto in alcuni suoi lavori precedenti³⁴, la necessità di una maggiore interazione tra scienza e filosofia per una reale comprensione dei problemi filosofici. Egli sviluppa una teoria generale della razionalità che costituisce un apprezzabile tentativo di conciliazione tra spiegazione filosofica e spiegazione scientifica dei fenomeni intenzionali e nella quale alle descrizioni di tipo logico-concettuale, caratteristiche della tradizione analitica della filosofia, si affiancano le descrizioni di tipo empirico caratteristiche delle ricerche neurobiologiche sulla natura e la configurazione della coscienza e del cervello umano.

³² Si veda il cap. 9 del presente volume.

³³ G. Edelman, *Il presente ricordato*, Rizzoli, Milano 1991; *Sulla materia della mente*, Adelphi, Milano 1993.

³⁴ J. Searle, *Il mistero della coscienza*, Raffaello Cortina, Milano 1998; *Mente, linguaggio, società*, Raffaello Cortina, Milano 2000.